

Superposition Enhanced Nested Sampling

Stefano Martiniani,^{*} Jacob D. Stevenson, David J. Wales, and Daan Frenkel

*University Chemical Laboratories, University of Cambridge,
Lensfield Road, Cambridge CB2 1EW, United Kingdom*

(Received 26 February 2014; revised manuscript received 3 June 2014; published 29 August 2014)

The theoretical analysis of many problems in physics, astronomy, and applied mathematics requires an efficient numerical exploration of multimodal parameter spaces that exhibit broken ergodicity. Monte Carlo methods are widely used to deal with these classes of problems, but such simulations suffer from a ubiquitous sampling problem: The probability of sampling a particular state is proportional to its entropic weight. Devising an algorithm capable of sampling efficiently the full phase space is a long-standing problem. Here, we report a new hybrid method for the exploration of multimodal parameter spaces exhibiting broken ergodicity. Superposition enhanced nested sampling combines the strengths of global optimization with the unbiased or athermal sampling of nested sampling, greatly enhancing its efficiency with no additional parameters. We report extensive tests of this new approach for atomic clusters that are known to have energy landscapes for which conventional sampling schemes suffer from broken ergodicity. We also introduce a novel parallelization algorithm for nested sampling.

DOI: [10.1103/PhysRevX.4.031034](https://doi.org/10.1103/PhysRevX.4.031034)

Subject Areas: Chemical Physics,
Computational Physics,
Statistical Physics

I. INTRODUCTION

Computer simulations play an important role in the study of phase transitions and critical phenomena. In particular, stochastic techniques such as Monte Carlo (MC) methods have proved to be powerful tools [1]. These methods rely on the ability of the Monte Carlo algorithm to sample the accessible volume in phase space. There are, however, many situations where standard Monte Carlo simulations suffer from a lack of ergodicity. In that case, more sophisticated algorithms are needed to explore the volume in phase space that is, in principle, accessible. Some such techniques are based on the efficient exploration of the underlying, multi-dimensional potential energy surface (PES) [2]. The PES, or energy landscape, can be viewed as a collection of basins separated by barriers, where each basin corresponds to a particular local minimum-energy configuration. The basin volumes define the entropic weight of the corresponding local minima. The transition rate from one basin to another depends on the barrier height as well as the relative entropic weights (configurational space volumes) [2]. Many PESs of interest exhibit frustration in the form of low-lying minima with different morphologies separated by high barriers. These structures may act as kinetic traps, when fixed-temperature

sampling methods such as molecular dynamics or Metropolis Monte Carlo sampling are used. There exists a wide range of extended or biased sampling techniques, both in Monte Carlo sampling and in molecular dynamics, that make it possible to speed up the sampling of landscapes with kinetic traps. These techniques include Monte Carlo methods, such as umbrella sampling [3,4]; density of states–based methods, such as the Wang-Landau method [5]; and replica exchange methods [6,7], along with their molecular-dynamics counterparts. Examples are the replica-exchange molecular dynamics method [8] and the metadynamics method [9]. In cases where a biased distribution is generated, the original distribution can be reconstructed using reweighting techniques [10,11]. However, these approaches may perform poorly when dealing with high-dimensional spaces exhibiting broken ergodicity or, in other words, with highly multimodal (or multifunnel [2,12–14]) parameter spaces [15–18].

In recent years, a Bayesian method known as nested sampling [19] has emerged as a possible alternative to extended or biased sampling methods. The nested sampling approach has found widespread applications in astrophysics [20,21] and cosmology [22,23] and has drawn the attention of computational and statistical physicists [24–29]. Furthermore, the method has recently been adopted for Bayesian model comparison in systems biology [30–32]. Nested sampling explores phase space in an unbiased way and allows one to determine statistically the density of states associated with shrinking fractions of phase space. This objective is achieved by placing a constraint on the potential energy (for instance), which decreases at each nested sampling iteration. Like Wang-Landau sampling, the

^{*}Corresponding author.
sm958@cam.ac.uk

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

method is athermal and produces the density of states and the partition function (Bayesian evidence) as its primary product. However, nested sampling does not require binning of the energy for systems with continuous potentials. The self-adapting steps in energy (but constant in log phase space volume) are attractive because the approach does not require prior knowledge of possible phase transitions. For example, the step size adjusts automatically as the phase space volume shrinks near a first-order phase transition [19,25].

An important drawback of nested sampling is that when the decreasing energy constraint forbids a transition to an unexplored basin, that basin cannot be visited and ergodicity is broken. Hence, while nested sampling certainly is conceptually interesting, its performance is often no better than that of conventional extended sampling methods in dealing with systems exhibiting broken ergodicity [25]. In the present work, we introduce a novel hybrid methodology for the exploration and thermodynamic analysis of such systems.

Superposition enhanced nested sampling (SENS) combines the strengths of unbiased global optimization techniques [2] with those of nested sampling. Global optimization techniques such as basin-hopping [33–35] are designed to find the lowest-energy configuration of a PES. They are not constrained to sample according to any distribution, so they are free to use “quick-and-dirty” techniques while searching for the global minimum. For example, they can take Monte Carlo steps that do not satisfy detailed balance and make use of minimization algorithms such as Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm and the conjugate gradient. Such operational freedom makes global optimization algorithms much more efficient than generalized ensemble methods at locating the lowest-energy minima [36–40]. Collections of the lowest-energy-minima configurations thus obtained can then be used in the context of the superposition approach [2,39–45] to compute the thermodynamic properties of the system. However, doing so accurately at high temperatures using the superposition approach alone often requires a prohibitively large number of minima.

In the present contribution, we show how knowledge of the lowest-energy minima and their statistical weights, calculated using the harmonic superposition approximation (HSA), can be exploited to enhance the problematic low-energy behavior of nested sampling, thus making it likely that none of the important minima and associated regions are missed. Although we discuss SENS in the context of energy landscapes, the method is completely general and can be applied to any multimodal parameter space whose minima (maxima in likelihood) can be identified by global optimization algorithms.

II. NESTED SAMPLING

Nested sampling [19] provides an elegant solution to the problem of evaluating the density of states, and hence the

partition function, for arbitrary systems. A likelihood value is assigned to each possible configuration. For our purposes, the likelihood is the Boltzmann factor $\exp(-E/k_B T)$, but it could be some other measure. Typically, there are large numbers of configurations with a low likelihood. In addition, there may be a small number of configurations with high likelihood.

The aim of nested sampling is to sample configuration space uniformly but with the energy constrained to lie below a maximum value E^{\max} that decreases iteratively throughout the calculation. The rate of decrease is maintained self-consistently, such that the phase space volume with energy less than E^{\max} decreases by a constant factor in each iteration.

The nested sampling algorithm starts by generating K configurations of the system completely at random, distributed uniformly, in configuration space. The energy $E_{\mathcal{R}}$ of each of these configurations is computed and added to a sorted list, where \mathcal{R} is the associated index in the sorted list. For each of these replicas, we define the configurational phase space volume $\Omega_{E \leq E_{\mathcal{R}}}$ containing all configurations with $E \leq E_{\mathcal{R}}$. The key insight of nested sampling is that the volumes $\Omega_{E \leq E_{\mathcal{R}}}$ normalized by the total phase space volume are distributed according to the beta distribution $\text{Beta}(K - \mathcal{R} + 1, \mathcal{R})$ [46]. This distribution has the following expectation value and variance:

$$\mu_{\mathcal{R}} = 1 - \frac{\mathcal{R}}{K + 1} \quad \text{and} \quad \sigma_{\mathcal{R}}^2 = \frac{\mathcal{R}(K - \mathcal{R} + 1)}{(K + 2)(K + 1)^2}. \quad (1)$$

The above formalism assumes that the total phase space volume Ω_{tot} is finite, but this condition can generally be satisfied with negligible error, for example, by placing the system in a large box.

At the i th nested sampling iteration, the replica (out of K replicas) with highest energy E_i^{\max} is discarded and replaced by a new configuration sampled uniformly under the constraint $E \leq E_i^{\max}$. The maximum energy E_i^{\max} is stored for later analysis. Again, the volume of configuration space with energy less than the \mathcal{R} th largest energy $\Omega_{E \leq E_{\mathcal{R}}}$, this time normalized by $\Omega_{E \leq E_i^{\max}}$, is distributed according to the beta distribution, with mean and variance given by Eq. (1). During the nested sampling iteration, the volume of phase space with energy below E^{\max} contracts, on average, by $\mu_1 = K/(K + 1)$. After N nested sampling iterations, the algorithm produces a list of the form $\{E_1^{\max}, E_2^{\max}, \dots, E_N^{\max}\}$. We can associate a fraction of configuration space $X_i = \Omega_{E \leq E_i^{\max}}/\Omega_{\text{tot}} = \mu_1^i$ with each E_i^{\max} . The density of states, or the (normalized) volume of phase space with energy between E_{i+1}^{\max} and E_i^{\max} , is simply

$$g_i(E) = X_i - X_{i+1} = \mu_1^i - \mu_1^{i+1} = \frac{1}{K + 1} \left(\frac{K}{K + 1} \right)^i. \quad (2)$$

Thermodynamic quantities of interest, such as the mean energy, entropy, free energy, and heat capacity, can easily

be computed from the density of states at an arbitrary temperature.

To generate configurations uniformly in space, we use the strategy suggested by Skilling [19]: After removing the configuration with the highest energy, one of the remaining $K - 1$ replicas (chosen randomly) is duplicated. The new configuration is then evolved through a Markov chain Monte Carlo (MCMC) walk sufficiently long to decorrelate the system from its initial state. This Monte Carlo walk is equivalent to a normal Monte Carlo simulation at infinite temperature. The coordinates are randomly perturbed, and the new configuration is accepted, subject only to the condition that the energy remains below E^{\max} . For most systems of interest, the vast majority of the computational effort will be spent generating new configurations.

A. Parallelization

Nested sampling can be formulated to run in parallel on an arbitrary number of processors. We present a pseudocode description of our parallel implementation in Algorithm 2. Since this scheme also constitutes the basic framework for SENS, we define the MCMC loop in the most general way at line 9 of Algorithm 2. For the purpose of discussing the algorithm in its simplest form, here, we will consider Algorithm 1.

```

For  $l = 0$  to  $N$  steps, do
  generate trial configuration (e.g., by random
  perturbation);
  if  $E_{\text{trial}} \leq E_{\text{max}}$ , accept trial configuration;
end for

```

ALGORITHM 1. Nested sampling MCMC.

At each nested sampling iteration, instead of removing only the replica with the highest energy, we remove the \mathcal{P} replicas with highest energy, where \mathcal{P} is the number of processors available. The rate of phase space contraction is now given by $\mu_{\mathcal{P}}$, leading to much faster phase space contraction and shorter calculations in terms of wall-clock time. This parallelization procedure was first described in Ref. [27]. Our improvement is that we do not discard the $\mathcal{P} - 1$ replicas with highest energy but we store them for later analysis. Phase space contraction between iterations is still constant, but now, the post analysis is slightly more complicated. The fraction of configuration space associated with the n th recorded energy is

$$X_n = \prod_{i=0}^n \frac{K - i \% \mathcal{P}}{K + 1 - i \% \mathcal{P}}, \quad (3)$$

where $\%$ is the mod operator. This method follows the same stepping routine as the existing parallelization algorithm. However, it produces \mathcal{P} times as many points, hence providing a more detailed picture of the potential energy surface and a much more fine-grained binning of the density of states.

III. SENS: THE CONCEPT

Global optimization is a common numerical problem, and global optimization algorithms have been developed in many areas of science [2,47,48]. Knowledge of the local minima alone, however, is not sufficient to infer all the observable properties of interest from the energy landscape (or, in general, any parameter space). The HSA [49] (for more details, see, e.g., Ref. [2]) allows one to compute the density of states and the partition function, solely based on the knowledge of the individual local minima and the local curvatures (normal-mode frequencies) of the potential energy landscape, via the Hessian matrix. In the HSA, each local minimum corresponds to a harmonic basin and observable properties are expressed as a sum over individual contributions of the minima.

▷ Initialization

- 1: Generate K random configurations;
 - 2: store their coordinates and energy in `LiveList`;
- ▷ main loop
- 3: while the termination condition is false, do
 - 4: remove the \mathcal{P} replicas $\{\mathbf{R}_m^{(1)}, \dots, \mathbf{R}_m^{(\mathcal{P})}\} \equiv \{\mathbf{R}_m\}$ with highest energy $\{E_m^{(1)} > \dots > E_m^{(\mathcal{P})}\} \equiv \{E_m\}$ from `LiveList`;
 - 5: append $\{E_m\}$ to `OutputList`;
 - 6: set $E_{\max} = E_m^{(\mathcal{P})}$;
 - 7: select \mathcal{P} replicas $\{\mathbf{R}_s^{(1)}, \dots, \mathbf{R}_s^{(\mathcal{P})}\} \equiv \{\mathbf{R}_s\}$ from `LiveList` at random;
 - 8: add a copy of $\{\mathbf{R}_s\}$ to `LiveList`;
 - 9: `MCMCLOOP`($\{\mathbf{R}_s\}, E_{\max}, \text{minima.db}$);
 - 10: end while

ALGORITHM 2. Parallel nested sampling.

The HSA has been shown to be very effective for several systems [18,40,50], but the accuracy depends on how well the potential energy of the basins can be approximated as harmonic and how many minima are thermodynamically important. While the HSA captures landscape anharmonicity, arising from the distribution of local minima, it does not include well anharmonicity, arising from the shape of the well. Therefore, the HSA becomes an increasingly good approximation at lower energies where well anharmonicity is less important. The total number of minima increases exponentially with system size [49,51], but it is impossible to tell *a priori* how many of those are important. For example, LJ₃₁, a cluster of 31 isotropic particles interacting through a Lennard-Jones potential [52], has about 3×10^{15} distinct minima [12], but only a few dozen are required to reproduce the low-temperature thermodynamic behavior.

The global resolution of nested sampling depends on the number of replicas K used in the simulation, which is generally limited by the available computation time (the larger the K , the slower the descent in energy). A more serious problem for nested sampling is that if the barrier to enter an unexplored funnel or superbasis is higher than the

energy constraint E_{\max} , that region of the PES will never be explored if it is not already populated. For example, in a crystallization transition, at high energy, the statistical weight of the liquid phase will be overwhelming and there will be no replicas in the region corresponding to the solid phase. However, as the energy constraint decreases (and hence the temperature), the relative statistical weight associated with the solid phase increases. If we could sample phase space uniformly, then at low energy, we would observe a phase transition corresponding to crystallization, but we must resort to a MCMC walk to explore phase space. If the entrance to the superbasin corresponding to the crystal has been locked out by E_{\max} , a Markov chain will not be able to find it, thus missing the transition.

Here, we propose a new method that combines complementary techniques: Nested sampling can sample efficiently the high-energy regions of phase space, while at low energy, a database of minima obtained by global optimization is used to augment the survey. While nested sampling assigns the correct statistical weight to each basin, global optimization makes it likely that no important minima are missed. This philosophy is also used in other methods combining replica exchange Monte Carlo sampling with global optimization algorithms to treat broken ergodicity [12,16,53,54].

IV. SENS: THE ALGORITHM

Employing knowledge of low-lying minima fits naturally within the framework of nested sampling. We present here both an exact and an approximate implementation of the SENS algorithm. Exact SENS is fully unbiased and requires no additional parameters than those needed in nested sampling. Approximate SENS, on the other hand, is formally biased and requires additional parameters. The reason for presenting both methods is that, in some cases, the latter approach can be considerably simpler to implement than the former, while generally producing equally good, or better, results. SENS is based on the original nested sampling algorithm presented in Algorithm 2. The novelty of our method resides in the augmented sampling of the parameter space obtained by coupling the MCMC walk to the HSA. SENS can therefore be implemented by changing the function $\text{MCLOOP}(\{\mathbf{R}_s\}, E_{\max}, \text{minima.db})$ of Algorithm 2. A full outline of the SENS algorithm can be found in Algorithm 3 of the Supplemental Material [55]. To run SENS, a database of the lowest-energy minima must be precomputed.

In this work, we adopt basin-hopping [33–35] as the global optimization algorithm of choice. Basin-hopping associates any given point of the PES to a minimum obtained by energy minimization, thus transforming the PES into a set of catchment basins. This basin transformation is combined with a sampling scheme to search for the global minimum. At each step, the coordinates of the current minimum configuration are perturbed to hop out the basin and minimized again to find a new minimum. Each step between two minima configurations is accepted with probability

$$P(\mathbf{x}_{\text{old}} \rightarrow \mathbf{x}_{\text{new}}) = \min\{1, \exp[-\beta(E_{\text{new}} - E_{\text{old}})]\}.$$

If the move is rejected, the coordinates are reset to those of the current local minimum. Since perturbations should be large enough to step out of the catchment basin, the step size is typically much larger than for thermodynamic sampling. Furthermore, since detailed balance need not hold, the step size can be dynamically adjusted to improve sampling. Basin-hopping has been successfully applied to a wide range of atomic, molecular, soft, and condensed matter systems [50,56–58].

A. Exact SENS

An unbiased version of SENS can be implemented by means of Hamiltonian replica exchange Monte Carlo moves [59,60]: In addition to normal MC steps, we introduce a new Monte Carlo step in which a minimum is sampled from the database according to its HSA configurational entropic weight:

$$w_c^{(b)}(E) = \frac{\Omega_c^{(b)}(E)}{\Omega_c(E)}. \quad (4)$$

We define the configurational volume of basin b

$$\Omega_c^{(b)}(E) \propto \frac{n_b(E - V^{(b)})^{\kappa/2}}{\prod_{\alpha=1}^{\kappa} \nu_{\alpha}^{(b)}} \quad (5)$$

and the total configurational volume

$$\Omega_c(E) \propto \sum_b \frac{n_b(E - V^{(b)})^{\kappa/2}}{\prod_{\alpha=1}^{\kappa} \nu_{\alpha}^{(b)}}, \quad (6)$$

where $V^{(b)}$ is the potential energy of the minimum corresponding to basin b , $\nu_{\alpha}^{(b)}$ are the normal-mode vibrational frequencies defined by the Hessian matrix, κ is the number of vibrational degrees of freedom, and n_b is the degeneracy of the basin. (For Lennard-Jones clusters, n_b is the number of distinct nonsuperimposable permutation-inversion isomers for minimum b [2].) Here, we leave out all the constant factors that cancel out as well as overall rotations. Once a minimum is selected, a configuration with $E \leq E_{\max}$ is then chosen uniformly from within its basin of attraction. This approach corresponds to selecting a point uniformly from a multidimensional harmonic well. Such a configuration can be generated analytically; see the Supplemental Material for details [55]. Unlike Ref. [16], in our approach, we sample from the uniform distribution of configurations below energy E_{\max} , rather than from the corresponding canonical distribution.

Thus, we obtain a configuration \mathbf{R}_{sys} sampled according to the true Hamiltonian \mathbb{H}_{sys} and a trial configuration \mathbf{R}_{har} sampled according to the HSA Hamiltonian \mathbb{H}_{har} . The energies of the two configurations are then computed with the other Hamiltonian. If

$$\mathbb{H}_{\text{har}}(\mathbf{R}_{\text{sys}}) \leq E_{\max} \quad \text{and} \quad \mathbb{H}_{\text{sys}}(\mathbf{R}_{\text{har}}) \leq E_{\max}, \quad (7)$$

then the true distribution and the HSA distributions overlap, the swap is accepted, and \mathbf{R}_{sys} becomes \mathbf{R}_{har} ; otherwise, it is rejected. This procedure guarantees that detailed balance is satisfied; for further discussion, refer to the Supplemental Material [55]. In practice, only the lowest-energy minima will successfully swap, since the HSA can only be reasonably accurate around these basins. It is, however, at low energy that such swaps are needed the most due to the hard-energy constraint used by nested sampling. Note that swaps are complemented by regular MCMC walks to allow for the exploration of the full configuration space. In SENS, the replicas are allowed to “tunnel” between basins, thus improving the sampling. A more detailed description, along with a pseudocode implementation of MCLOOP specific to Lennard-Jones clusters, is provided in the Supplemental Material [55].

B. Approximate SENS

The implementation of approximate SENS is somewhat simpler but comes at the cost of at least one extra parameter. The basic idea of approximate SENS is that the sampling of configuration space can be augmented by starting a MCMC walk from a local minimum configuration, sampled from the database according to its entropic weight [Eq. (4)], with some user-defined frequency. This frequency is intrinsically defined in exact SENS by the relative overlap of the HSA and the true density of states. To implement approximate SENS, we only need a database of minima and their relative weights computed according to Eq. (4). Before each MCMC step, a random number is drawn. If this number is less than some user-defined probability P_{DS} , then a minimum is selected from the database according to the HSA weights and the MCMC walk starts from this minimum configuration. A pseudocode implementation of MCLOOP for approximate SENS is provided in the Supplemental Material [55].

There are two main sources of bias in the approximate SENS. The first one is due to the limited number of minima from which we sample, since we cannot include the large number of high-energy minima. The second source of error is due to the poor quality of the HSA approximation far from the minimum; hence, the entropic weights for the minima are not accurate at high energy. The most obvious way of reducing these biases is to use long MCMC walks. In fact, if we sample from the wrong basin, a long MCMC walk will allow the system to escape and explore regions of phase space with greater entropic weight. However, very long MCMC walks are computationally expensive, and if short runs are required, we need to sample from the database of minima carefully. If we start sampling from the database of minima at high energy, we will possibly introduce a bias due to overweighting of the low-energy regions of configuration space. To avoid this problem, we suppress sampling from the database until we are sure the HSA is likely to describe the potential energy landscape

accurately. We use a simple function (of the Fermi type) that delays the onset of sampling from the database of minima and limits its maximum frequency

$$f_{\text{onset}} = \frac{f_{\text{max}}}{1 + e^{(E_{\text{min}}^{(\mathcal{R})} - E_{\text{on}})/\Delta E}}, \quad (8)$$

where E_{on} is some onset energy and $E_{\text{min}}^{(\mathcal{R})}$ is the energy of the replica with lowest energy. E_{on} could be chosen as $E_{\text{max}}^{(\text{minima.db})}$, the energy of the highest known minimum (stored in the database), or as the largest energy at which the HSA describes the system accurately. f_{max} and ΔE are user-defined parameters that determine the total probability of sampling after the onset and the width of the onset region, respectively. For small sampling probabilities $P_{\text{DS}} \ll 1$, the optimal frequency of sampling from the database should scale as $1/K$; a theoretical justification is derived in the Supplemental Material [55]. Hence, for $P_{\text{DS}} \ll 1$, we can make the probability of sampling from the database independent of the number of replicas, replacing f_{max} with f_{max}/K .

We identify two possible strategies for the application of approximate SENS. One is to start sampling from a large database early in the simulation when $E_{\text{on}} = E_{\text{max}}^{(\text{minima.db})}$, with a small P_{DS} ; hence, we choose $f_{\text{max}} \ll 1$. This procedure allows nested sampling to do most of the work but ensures that no important basins will be missed. Alternatively, sampling from the database can be delayed until all the high-temperature transitions have occurred, at which point we start sampling more extensively from the database, hence $f_{\text{max}} \gtrsim 1/2$. Note that the database can be considerably smaller in this case. The first strategy is a slight enhancement to nested sampling, while the latter strategy interpolates between nested sampling and the HSA in a similar spirit to the basin-sampling method [12]. Importantly, even if we sample from the database of minima, we use the MCMC walk to explore more the anharmonic regions of a basin, allowing us to go beyond the harmonic approximation.

V. RESULTS

We test SENS by calculating the thermodynamic properties of Lennard-Jones clusters exhibiting broken ergodicity. LJ clusters are systems of particles that interact via the Lennard-Jones potential [52]

$$E = 4\epsilon \sum_{i < j} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right], \quad (9)$$

where ϵ is the pair well depth, σ is the separation at which $E = 0$, and $2^{1/6}\sigma$ is the equilibrium pair separation. LJ clusters have served as benchmarks for many global optimization techniques and thermodynamics sampling [2,12,15,18,25,34].

The majority of putative ground states for LJ clusters are based on icosahedral packings [13]. For some magic

number LJ clusters complete Mackay icosahedra are possible, for example, $N = 13, 55$. Complete icosahedral structures are considerably more stable than neighboring sizes, and their landscape is funneled towards the global minimum [13]. There are, however, other sizes for which the global minimum is not icosahedral. Examples are LJ₃₈, whose ground state is an fcc-truncated octahedron [13], and LJ₇₅, whose global minimum is a Marks decahedron [13]. Because of the overwhelming number (entropic weight) of structures based on incomplete icosahedra at high energy, the energy landscapes of LJ clusters with nonicosahedral global minima exhibit broken ergodicity. Calculating accurate thermodynamic properties for these systems has proved to be a real challenge for all conventional techniques [2,15,18,25], and hybrid or more complicated schemes [12,14,15,18] are necessary. LJ clusters with broken ergodicity therefore provide excellent benchmarks to test the performance of new sampling techniques.

A. LJ₃₁

LJ₃₁ is the smallest Lennard-Jones cluster exhibiting broken ergodicity and a low-temperature solid-solid phase-like transition from Mackay to anti-Mackay surface structures [13]. Convergence of the heat capacity curve for LJ₃₁ by parallel tempering (PT) with 24 geometrically distributed temperatures in the range 0.0125 to 0.6 required $N_E^{\text{total}} = 1.9 \times 10^{11}$ energy evaluations to converge (curve shown in Fig. 1). Pártay *et al.* [25] report that $K = 288\,000$ replicas and $N_E^{\text{total}} = 3.4 \times 10^{12}$ energy evaluations were needed to converge the heat capacity curve of LJ₃₁ by nested sampling (NS) using a low particle density of $2.31 \times 10^{-3} \sigma^{-3}$ (100-fold less dense than our system). Figure 1 compares the heat capacity curves obtained by

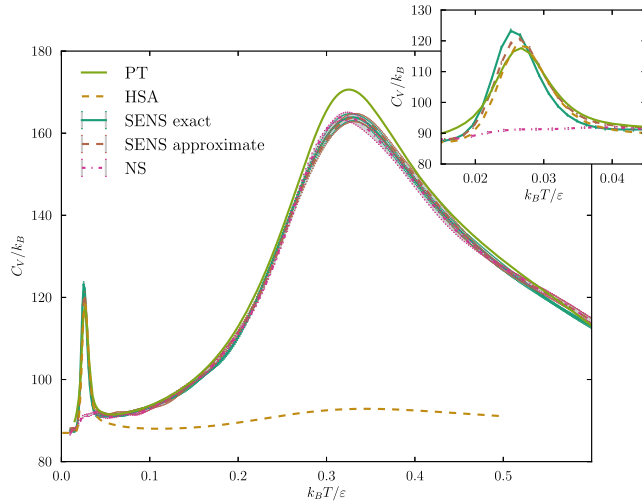


FIG. 1. Heat capacity curves for LJ₃₁. PT and HSA correspond to parallel tempering and the harmonic superposition approximation, respectively. All SENS calculations were performed using $K = 20\,000$ replicas.

PT, HSA (computed using $\geq 80\,000$ minima), NS, and SENS for LJ₃₁. The SENS and NS results correspond to $K = 20\,000$ replicas, $N = 10\,000$ steps for each MCMC walk, and $\mathcal{P} = 16$ cores. The database of minima used for SENS contained the lowest 183 minima, although for exact SENS, we observe that only seven minima contribute to the swaps; see Table IV of the Supplemental Material for the swap statistics [55]. From Fig. 1, we see that both exact SENS and approximate SENS are well converged and agree with the PT curve over the whole temperature range and with the HSA at low temperature. We note that $K = 20\,000$ replicas are not nearly enough for NS to converge, and the low-temperature peak is, in fact, completely absent. Using this number of replicas, SENS requires half the total number of energy evaluations of PT and 1 order of magnitude less than NS; see Table I. The swap operations do not constitute a noticeable overhead, and the reduction in the total number of energy evaluations corresponds to an equivalent reduction in wall-clock time.

In Fig. 2, we show a comparison of PT, HSA, and exact SENS for a range of replica numbers $2500 \leq K \leq 20\,000$; see Table I for comparison. We observe that the high-temperature peak practically converges for $K = 10\,000$, and it resembles the features of the converged curve quite well, even for smaller numbers of replicas. The low-temperature peak instead converges very quickly, for as few as $K = 2500$ replicas, representing an improvement in performance of 20 times over PT. We note that one of the great strengths of SENS is that even when a small number of replicas are used and run times are very short, although the curves may not be completely converged, the physical picture produced by the method is always correct because all the important basins are visited. On the other hand, rapid convergence of the heat capacity curves requires the HSA to be a good representation for the system. LJ₃₈ is an

TABLE I. Comparison of methods used to obtain the LJ₃₁ heat capacity curves shown in Figs. 1 and 2. $N_E^{(\text{total})}$ indicates the total number of energy evaluations (summed over all processors). PT was performed using 24 replicas spread geometrically through the temperature range 0.0125 to 0.6. Note that approximate SENS can perform as well as exact SENS when fewer replicas are used; in the interest of brevity, we do not include these results, as the LJ₇₅ calculations illustrate clearly the capabilities of the method.

LJ ₃₁			
Method	K	N	$N_E^{(\text{total})}$
PT			1.9×10^{11}
NS (Ref. [25])	280 000		3.4×10^{12}
NS	20 000	10 000	1×10^{11}
SENS approximate	20 000	10 000	1×10^{11}
SENS exact	20 000	10 000	1×10^{11}
SENS exact	10 000	10 000	5.2×10^{10}
SENS exact	5000	10 000	2.6×10^{10}
SENS exact	2500	10 000	1.3×10^{10}

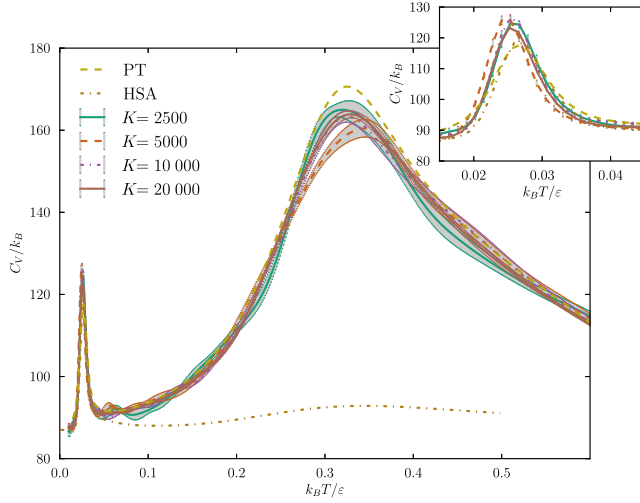


FIG. 2. Comparison of heat capacity curves for LJ₃₁ obtained by exact SENS using different numbers of replicas. The PT and HSA curves were obtained by parallel tempering and the harmonic superposition approximation, respectively.

example for which this condition does not hold as well; see the Supplemental Material for further details [55].

B. LJ₇₅

LJ₇₅ is a particularly clear example of a double-funneled energy landscape [13] with $O(10^{25})$ distinct local minima [12]. The decahedral global minimum is separated by a very large potential energy barrier from the lowest icosahedral minimum. Sharapov and Mandelshtam [17] showed that $O(10^{12})$ (total) energy evaluations of adaptive parallel tempering are not enough to converge the heat capacity peak corresponding to the solid-solid phaselike transition in LJ₇₅ [16]. Furthermore, the rate of convergence slows down dramatically (it practically stops) after $O(10^{11})$ (total) energy evaluations, and coupling of PT to the HSA is necessary to obtain convergence of the low-temperature peak [16]. Figure 3 compares the heat capacity curves obtained by HSA (computed using 758 minima) and SENS for LJ₇₅. SENS was carried out using $K = 30\,000$ or $K = 60\,000$ replicas and $N = 10\,000$ steps for each MCMC walk on $\mathcal{P} = 16$ processors. The database of minima for SENS contained the lowest 758 minima. Approximate SENS started sampling from the database at $E_{\text{on}} = -369\epsilon$, while for exact SENS, only ten of the minima contribute to the swaps; see Table VI of the Supplemental Material for swap statistics [55]. Unlike adaptive PT [16], approximate SENS converges in $O(10^{11})$ energy evaluations (Table II), but exact SENS fails to converge the low-temperature peak for the same number of replicas. As for LJ₃₈, exact SENS does not converge quickly due to the lower accuracy of the HSA, as inferred from the extremely low swap acceptance (Table VI of the Supplemental Material [55]). On the other hand, approximate SENS performs considerably better than for LJ₃₈ because the melting transition is well separated

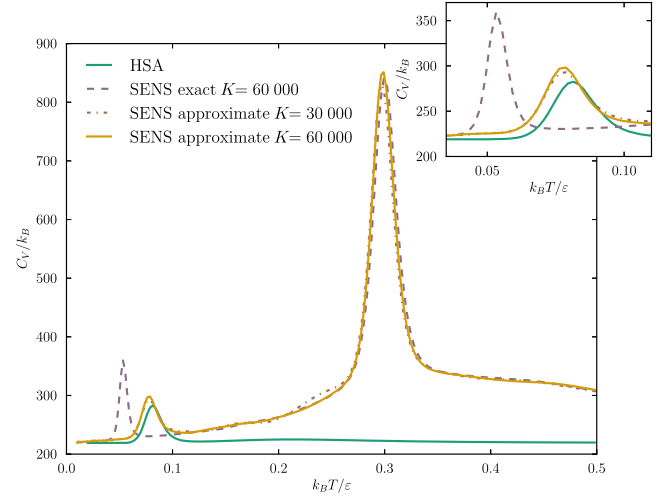


FIG. 3. Heat capacity curves for LJ₇₅. The PT and HSA results were obtained by parallel tempering and the harmonic superposition approximation, respectively. Exact SENS calculations were performed using $K = 60\,000$ replicas, while results for approximate SENS calculations are shown for both $K = 30\,000$ and $K = 60\,000$ replicas.

from the solid-solid transition, thus allowing sampling from the database relatively early on in the simulation (right after melting) without affecting the melting transition.

C. Methods

We define a move in a MCMC walk as the displacement of each individual particle along a random vector (n in total). After each MCMC walk, we update the step size in order to keep the average acceptance ratio within range of some target value, which we have chosen as 0.5. The default parameter values for the onset function [Eq. (8)] are $f_{\text{max}} \approx 2/3$ and $\Delta E = 1$. We used a spherical box of radius $R = 2.5\sigma$ for $n = 31$, $R = 2.8\sigma$ for $n = 38$, and $R = 3.0\sigma$ for $n = 75$, with no periodic boundary conditions and no cutoff radius. All calculations were carried out on a single workstation with $\mathcal{P} = 16$ processors (eight-core dual Xeon

TABLE II. Comparison of methods used to obtain the LJ₇₅ heat capacity curves shown in Fig. 3. $N_E^{(\text{total})}$ indicates the total number of energy evaluations (summed over all processors). PT curves are not shown, as the computational cost to converge its heat capacity by this method is computationally prohibitive, as shown in Ref. [16]. Exact SENS does not converge as quickly as approximate SENS due to the low accuracy of the HSA, hence the low swap acceptance.

LJ ₇₅			
Method	K	N	$N_E^{(\text{total})}$
SENS approximate	30 000	10 000	4×10^{11}
SENS approximate	60 000	10 000	8×10^{11}
SENS exact	60 000	10 000	8×10^{11}

E5-2670 2.6 GHz, Westmere) using the improved parallelization scheme discussed in Sec. II A. The calculations were terminated when the energy difference between the replicas with highest and lowest energies was less than $10^{-2}e$. Energies of the final “live” replicas were added to the output, and the compression factor associated with the ℓ th live replica was computed as

$$\mu_\ell^{(\text{live})} = \prod_{j=0}^{\ell < K} \frac{K - j}{K - j + 1}. \quad (10)$$

Error bars were obtained by the compression-factor resampling scheme discussed in the Supplemental Material [55]. By nested sampling or SENS iterations N_{iter} , we mean a whole nested sampling iteration on \mathcal{P} processors; the total number of energy evaluations is $N_E^{(\text{tot})} = N \times \mathcal{P} \times N_{\text{iter}}$, where N is the number of steps in a MCMC walk. The computational overhead associated with global optimization by basin-hopping is insignificant, as fewer than around $O(10^5)$ energy evaluations are necessary to find the global minima of the LJ clusters considered here [35]. Highly modular Python/C parallel implementations of nested sampling and SENS are publicly available [61,62].

ACKNOWLEDGMENTS

We gratefully acknowledge discussions with Andy Ballard, Victor Rühle, Thomas Stecher, Tine Curk, Boris Fackovec, and Robert Baldock. This work has been supported by the European Research Council and the EPSRC. S. M. acknowledges financial support from the Gates Cambridge Scholarship and J. D. S. from the Marie Curie IEF 275544. D. F. acknowledges support from the ERC Advanced Grant No. 227758 and the Wolfson Merit Award 2007/R3 of the Royal Society of London.

S. M. and J. D. S. have contributed equally to this work.

-
- [1] D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, England, 2005).
- [2] D. J. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge University Press, Cambridge, England, 2005).
- [3] G. M. Torrie and J. P. Valleau, *Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling*, *J. Comput. Phys.* **23**, 187 (1977).
- [4] B. A. Berg and T. Neuhaus, *Multicanonical Ensemble: A New Approach to Simulate First-Order Phase Transitions*, *Phys. Rev. Lett.* **68**, 9 (1992).
- [5] F. Wang and D. P. Landau, *Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States*, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [6] R. H. Swendsen and J.-S. Wang, *Replica Monte Carlo Simulation of Spin-Glasses*, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [7] R. H. Swendsen and J.-S. Wang, *Nonuniversal Critical Dynamics in Monte Carlo Simulations*, *Phys. Rev. Lett.* **58**, 86 (1987).
- [8] Y. Sugita and Y. Okamoto, *Replica-Exchange Molecular Dynamics Method for Protein Folding*, *Chem. Phys. Lett.* **314**, 141 (1999).
- [9] A. Laio and M. Parrinello, *Escaping Free-Energy Minima*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).
- [10] A. M. Ferrenberg and R. H. Swendsen, *New Monte Carlo Technique for Studying Phase Transitions*, *Phys. Rev. Lett.* **61**, 2635 (1988).
- [11] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, *Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations*, *J. Chem. Theory Comput.* **3**, 26 (2007).
- [12] D. J. Wales, *Surveying a Complex Potential Energy Landscape: Overcoming Broken Ergodicity Using Basin-Sampling*, *Chem. Phys. Lett.* **584**, 1 (2013).
- [13] 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 (1999).
- [14] J. P. K. Doye and D. J. Wales, *Thermodynamics of Global Optimization*, *Phys. Rev. Lett.* **80**, 1357 (1998).
- [15] P. Poulain, F. Calvo, R. Antoine, M. Broyer, and Ph. Dugourd, *Performances of Wang-Landau Algorithms for Continuous Systems*, *Phys. Rev. E* **73**, 056704 (2006).
- [16] V. A. Mandelshtam, P. A. Frantsuzov, and F. Calvo, *Structural Transitions and Melting in LJ74-78 Lennard-Jones Clusters from Adaptive Exchange Monte Carlo Simulations*, *J. Phys. Chem. A* **110**, 5326 (2006).
- [17] 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 (2007).
- [18] V. A. Sharapov, D. Meluzzi, and V. A. Mandelshtam, *Low-Temperature Structural Transitions: Circumventing the Broken-Ergodicity Problem*, *Phys. Rev. Lett.* **98**, 105701 (2007).
- [19] J. Skilling, *Nested Sampling for General Bayesian Computation*, *Bayesian Anal.* **1**, 833 (2006).
- [20] P. Mukherjee, D. Parkinson, and A. R. Liddle, *A Nested Sampling Algorithm for Cosmological Model Selection*, *Astrophys. J.* **638**, L51 (2006).
- [21] F. Feroz and M. P. Hobson, *Multimodal Nested Sampling: An Efficient and Robust Alternative to Markov Chain Monte Carlo Methods for Astronomical Data Analyses*, *Mon. Not. R. Astron. Soc.* **384**, 449 (2008).
- [22] J. R. Shaw, M. Bridges, and M. P. Hobson, *Efficient Bayesian Inference for Multimodal Problems in Cosmology*, *Mon. Not. R. Astron. Soc.* **378**, 1365 (2007).
- [23] F. Feroz, M. P. Hobson, and M. Bridges, *MultiNest: An Efficient and Robust Bayesian Inference Tool for Cosmology and Particle Physics*, *Mon. Not. R. Astron. Soc.* **398**, 1601 (2009).
- [24] I. Murray, D. J. C. MacKay, Z. Ghahramani, and J. Skilling, *Nested Sampling for Potts Models*, *Adv. Neural Inf. Process. Syst.* **18**, 947 (2006).
- [25] L. B. Pártay, A. P. Bartók, and G. Csányi, *Efficient Sampling of Atomic Configurational Spaces*, *J. Phys. Chem. B* **114**, 10502 (2010).

- [26] S. O. Nielsen, *Nested Sampling in the Canonical Ensemble: Direct Calculation of the Partition Function from NVT Trajectories*, *J. Chem. Phys.* **139**, 124104 (2013).
- [27] N. S. Burkoff, C. Várnai, S. A. Wells, and D. L. Wild, *Exploring the Energy Landscapes of Protein Folding Simulations with Bayesian Computation*, *Biophys. J.* **102**, 878 (2012).
- [28] L. B. Pártay, A. P. Bartók, and G. Csányi, *Nested Sampling for Materials: The Case of Hard Spheres*, *Phys. Rev. E* **89**, 022302 (2014).
- [29] B. J. Brewer, L. B. Pártay, and G. Csányi, *Diffusive Nested Sampling*, *Stat. Comput.* **21**, 649 (2011).
- [30] S. Aitken and O. Akman, *Nested Sampling for Parameter Inference in Systems Biology: Application to an Exemplar Circadian Model*, *BMC Syst. Biol.* **7**, 72 (2013).
- [31] R. Dybowski, T. J. McKinley, P. Mastroeni, and O. Restif, *Nested Sampling for Bayesian Model Comparison in the Context of Salmonella Disease Dynamics*, *PLoS One* **8**, e82317 (2013).
- [32] N. Pullen and R. J. Morris, *Bayesian Model Comparison and Parameter Inference in Systems Biology Using Nested Sampling*, *PLoS One* **9**, e88419 (2014).
- [33] Z. Li and H. A. Scheraga, *Monte Carlo-Minimization Approach to the Multiple-Minima Problem in Protein Folding*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6611 (1987).
- [34] D. J. Wales and J. P. K. Doye, *Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms*, *J. Phys. Chem. A* **101**, 5111 (1997).
- [35] M. T. Oakley, R. L. Johnston, and D. J. Wales, *Symmetrisation Schemes for Global Optimisation of Atomic Clusters*, *Phys. Chem. Chem. Phys.* **15**, 3965 (2013).
- [36] D. J. McGinty, *Vapor Phase Homogenous Nucleation and the Thermodynamic Properties of Small Clusters of Argon Atoms*, *J. Chem. Phys.* **55**, 580 (1971).
- [37] J. J. Burton, *Vibrational Frequencies and Entropies of Small Clusters of Atoms*, *J. Chem. Phys.* **56**, 3133 (1972).
- [38] M. R. Hoare, *Structure and Dynamics of Simple Microclusters*, *Adv. Chem. Phys.* **40**, 49 (1979).
- [39] F. H. Stillinger and T. A. Weber, *Packing Structures and Transitions in Liquids and Solids*, *Science* **225**, 983 (1984).
- [40] B. Strodel and D. J. Wales, *Free Energy Surfaces from an Extended Harmonic Superposition Approach and Kinetics for Alanine Dipeptide*, *Chem. Phys. Lett.* **466**, 105 (2008).
- [41] D. J. Wales, *Coexistence in Small Inert Gas Clusters*, *Mol. Phys.* **78**, 151 (1993).
- [42] J. P. K. Doye and D. J. Wales, *An Order Parameter Approach to Coexistence in Atomic Clusters*, *J. Chem. Phys.* **102**, 9673 (1995).
- [43] J. P. K. Doye and D. J. Wales, *Calculation of Thermodynamic Properties of Small Lennard-Jones Clusters Incorporating Anharmonicity*, *J. Chem. Phys.* **102**, 9659 (1995).
- [44] F. Calvo, J. P. K. Doye, and D. J. Wales, *Characterization of Anharmonicities on Complex Potential Energy Surfaces: Perturbation Theory and Simulation*, *J. Chem. Phys.* **115**, 9627 (2001).
- [45] 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).
- [46] The normalized fraction of configuration space $X_i = \Omega_{E \leq E_R} / \Omega_{\text{tot}} \in (0, 1)$. Since nested sampling assumes uniform sampling in phase space, estimating X_i becomes analogous to estimating the \mathcal{R} th largest position out of K points sampled at random on a unit segment. The point with the \mathcal{R} th largest position will be distributed on the unit line as $\text{beta}(K - \mathcal{R} + 1, \mathcal{R})$.
- [47] B. Hartke, *Global Optimization*, *Comput. Mol. Sci.* **1**, 879 (2011).
- [48] D. J. Wales, *Global Optimization of Clusters, Crystals, and Biomolecules*, *Science* **285**, 1368 (1999).
- [49] F. H. Stillinger and T. A. Weber, *Packing Structures and Transitions in Liquids and Solids*, *Science* **225**, 983 (1984).
- [50] S. Somani and D. J. Wales, *Energy Landscapes and Global Thermodynamics for Alanine Peptides*, *J. Chem. Phys.* **139**, 121909 (2013).
- [51] D. J. Wales and J. P. K. Doye, *Stationary Points and Dynamics in High-Dimensional Systems*, *J. Chem. Phys.* **119**, 12409 (2003).
- [52] J. E. Lennard-Jones, *Cohesion*, *Proc. Phys. Soc. London* **43**, 461 (1931).
- [53] R. Zhou and B. J. Berne, *Smart Walking: A New Method for Boltzmann Sampling of Protein Conformations*, *J. Chem. Phys.* **107**, 9185 (1997).
- [54] I. Andricioaei, J. E. Straub, and A. F. Voter, *Smart Darting Monte Carlo*, *J. Chem. Phys.* **114**, 6994 (2001).
- [55] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.4.031034> for details of how to sample configurations from a harmonic well, discussion of detailed balance for the Hamiltonian replica-exchange moves, discussion of the scaling of the onset function in Eq. (8), complete pseudocode implementations of the algorithms described in the main text, results for LJ₃₈ cluster and swap statistics for the exact SENS calculations.
- [56] K. Mochizuki, C. S. Whittleston, S. Somani, H. Kusumaatmaja, and D. J. Wales, *A Conformational Factorisation Approach for Estimating the Binding Free Energies of Macromolecules*, *Phys. Chem. Chem. Phys.* **16**, 2842 (2014).
- [57] S. W. Olesen, S. N. Fejer, D. Chakrabarti, and D. J. Wales, *A Left-Handed Building Block Self-Assembles into Right- and Left-Handed Helices*, *R. Soc. Chem. Adv.* **3**, 12905 (2013).
- [58] C. J. Forman, S. N. Fejer, D. Chakrabarti, P. D. Barker, and D. J. Wales, *Local Frustration Determines Molecular and Macroscopic Helix Structures*, *J. Phys. Chem. B* **117**, 7918 (2013).
- [59] A. Bunker and B. Dünweg, *Parallel Excluded Volume Tempering for Polymer Melts*, *Phys. Rev. E* **63**, 016701 (2000).
- [60] H. Fukunishi, O. Watanabe, and S. Takada, *On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction*, *J. Chem. Phys.* **116**, 9058 (2002).
- [61] S. Martiniani and J. D. Stevenson, *Nested Sampling*, https://github.com/js850/nested_sampling.
- [62] S. Martiniani and J. D. Stevenson, *Superposition Enhanced Nested Sampling*, <https://github.com/smcantab/sens>.