(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 January 2013 (10.01.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/006804 A1

(51) **International Patent Classification**:
G06F 17/50 (2006.01)

(21) **International Application Number**:
PCT/US2012/045779

(22) **International Filing Date**:
6 July 2012 (06.07.2012)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/504,926 6 July 2011 (06.07.2011) US

(71) **Applicant** *(for all designated States except US)*: **BROWN UNIVERSITY** [US/US]; Box 1949, Providence, RI 02912 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **DUPUIS, Paul**; . (US). **DOLL, Jimmie, D.** [US/US]; 51 Bay Road, Barrington, RI 02806 (US).

(74) **Agents**: **SCOZZAFAVA, Mary, Rose** et al.; Wilmer Cutler Pickering Hale And Dorr LLP, 60 State Street, Boston, MA 02109 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: RARE EVENT SAMPLING

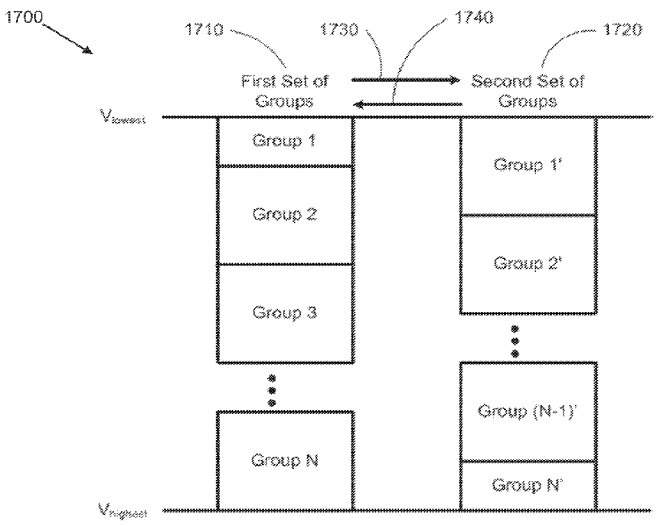

FIG. 17

(57) **Abstract**: Systems and methods are disclosed for a new approach to the rare-event Monte Carlo sampling problem. This technique utilizes a symmetrization strategy to create probability distributions that are more highly connected and thus more easily sampled than their original, potentially sparse counterparts. The methods can be applied to Lennard-Jones clusters of varying complexity and rare-event character. In one embodiment, a system is disclosed comprising a grouping module for dividing a control variable range into a first set of groups and a second set of groups, an analysis module for computing data representative of the physical system over the set of control values in each group in the first and second set of groups and for accelerating the flow of information by simultaneously exchanging information between all of the control values within a group; and a swapping module for communicating data.

WO 2013/006804 A1

RARE-EVENT SAMPLING

Cross-Reference to Related Patent Applications

[0001] This application claims the benefit of priority to U.S. Pat. App. No. 61/504,926, filed July 6, 2011, which is hereby incorporated by reference in its entirety.

Government Support

[0002] This research was partially supported by the DOE Multiscale and Optimization for Complex Systems Program (DESC0002413 and DE-00015561), the Army Research Office (W911NF-09-1-0155), the Air Force Office of Scientific Research (FA9550-07-1-0544, FA9550-09-1-0378), and by the National Science Foundation (DMS-I008331). The U.S. Government may have certain rights in this invention as provided for by the terms of the above grants.

Background

[0003] This invention relates to simulating physical properties of matter in a computer system. More particularly, the invention relates to increasing the speed at which useful simulations of physical systems can be performed. In conventional simulations, large amounts of computational resources are expended simulating a physical system in which nothing of interest is occurring. The invention increases the likelihood that events of interest in the physical system occur during a simulation and increases the efficiency with which computational resources are used.

[0004] Monte Carlo methods are among the more versatile and widely utilized tools in modem simulation. In particular, they provide a refinable means for treating problems of a physically realistic complexity. They are of importance in classical and quantum statistical-mechanical applications, for example, where the problem of computing thermodynamic properties can be reduced to that of performing well-defined, large-dimensional averages over known probability distributions. Such applications are the primary focus of the present disclosure.

Summary

[0005] Systems and methods are disclosed for a new approach to the rare-event Monte Carlo sampling problem. This technique utilizes a symmetrization strategy to create probability distributions that are more highly connected and thus more easily sampled than their original,

potentially sparse counterparts. The methods can be applied to many physical and statistical problems, such as Lennard-Jones clusters of varying complexity and rare-event character.

[0006] In one embodiment, a system is disclosed comprising a grouping module for dividing a control variable range into a first set of groups and a second set of groups, an analysis module for computing data representative of the physical system over the set of control values in each group in the first and second set of groups and for accelerating the flow of information by simultaneously exchanging information between all of the control values within a group; and a swapping module for communicating data.

[0007] The control variable range extends from a lowest value, V_{lowest}, to a highest value, $V_{highest}$, such that each group in the first set including a discrete set of control values that extend over only a portion of the range from V_{lowest} to $V_{highest}$, each group in the second set including a discrete set of control values that extend over only a portion of the range from V_{lowest} to $V_{highest}$, each group in the first set being distinct from each group in the second set, and the control values in the first set extend from V_{lowest} to $V_{highest}$, and the control values in the second set extend from V_{lowest} to $V_{highest}$.

[0008] The swapping module is such that analysis module can use data computed using a group in the first set for computing data representative of the physical system for a group of control values in the second set, and the analysis module can use data computed using a group in the second set for computing data representative of the physical system for a group of control values in the first set.

[0009] The description and figures which follow describe various embodiments in further detail. Although the present disclosure is described and illustrated in the following embodiments, it is understood that the present disclosure has been made only by way of example, and that numerous changes in the details of implementation of the disclosure may be made without departing from the spirit and scope of the disclosure.

Brief Description of the Drawings

[0010] FIG. 1 is a schematic functional plot showing temperature-swapped and particle-swapped processes, in accordance with some embodiments.

[0011] FIG. 1A is a schematic functional plot showing an exemplary potential energy function, in accordance with some embodiments.

[0012] FIG. 2 is a schematic functional plot showing partial infinite swapping handoffs, in accordance with some embodiments.

[0013] FIG. 2A is a multidimensional plot showing isosurfaces, in accordance with some embodiments.

[0014] FIG. 3 is a potential energy plot for a 13-atom Lennard-Jones cluster, in accordance with some embodiments.

[0015] FIG. 3A is a multidimensional plot showing symmetrized isosurfaces, in accordance with some embodiments.

[0016] FIG. 4 is a potential energy plot for a 13-atom Lennard-Jones cluster, in accordance with some embodiments.

[0017] FIG. 4A is a multidimensional plot showing partially-symmetrized isosurfaces, in accordance with some embodiments.

[0018] FIG. 4B is a second multidimensional plot showing partially-symmetrized isosurfaces, in accordance with some embodiments.

[0019] FIG. 5 is a potential energy plot for a 38-atom Lennard-Jones cluster showing parallel tempering and partial infinite swapping, in accordance with some embodiments.

[0020] FIG. 5A is a temperature profile plot for four relaxation simulations, in accordance with some embodiments.

[0021] FIG. 5B is a relaxation curve plot using infinite swapping, in accordance with some embodiments.

[0022] FIG. 6 is a cooling portion of the lowest-temperature curve in FIG. 5b, in accordance with some embodiments.

[0023] FIG. 7 is a plot of a representative value on the parallel tempering relaxation curves of FIG. 6, in accordance with some embodiments.

[0024] FIG. 8 is a plot of a representative value on the parallel tempering relaxation curves of FIG. 6 using partial infinite swapping, in accordance with some embodiments.

[0025] FIG. 9 is a loop average potential energy plot for a 38-atom Lennard-Jones cluster, in accordance with some embodiments.

[0026] FIG. 10 is a heat capacity plot for a 38-atom Lennard-Jones cluster, in accordance with some embodiments.

[0027] FIG. 11 is a loop average potential energy plot for a 38-atom Lennard-Jones cluster with an alternate simulation start point, in accordance with some embodiments.

[0028] FIG. 12 is a heat capacity plot for a 38-atom Lennard-Jones cluster with an alternate simulation start point, in accordance with some embodiments.

[0029] FIG. 13 is a relaxation plot for a 38-atom Lennard-Jones cluster, in accordance with some embodiments.

[0030] FIG. 14 is an equilibration plot for an LJ-31 system, in accordance with some embodiments.

[0031] FIG. 15 is a heat capacity plot for an LJ-31 system, in accordance with some embodiments.

[0032] FIG. 16 is a schematic diagram of the partial infinite swapping approach, in accordance with some embodiments.

[0033] FIG. 17 is a schematic diagram of a grouping method, in accordance with some embodiments.

[0034] FIG. 18 is a schematic diagram of a computer system, in accordance with some embodiments.

Detailed Description

[0035] Introduction

[0036] Monte Carlo methods are among the more versatile and widely utilized tools in modem simulation. In particular, they provide a refinable means for treating problems of a physically realistic complexity. They are of importance in classical and quantum statistical-mechanical applications, for example, where the problem of computing thermodynamic properties can be reduced to that of performing well-defined, large-dimensional averages over known probability distributions. Such applications are the primary focus of the present discussion.

[0037] A key step in the use of Monte Carlo methods in either a classical or quantum equilibrium statistical-mechanical context is generating a suitable sampling of the relevant probability distribution. While straightforward in principle, difficulties can arise in practice if the probability distribution involved is "sparse". In the case of sparse distributions the diffusive random walk methods generally used to perform the sampling can become problematic as transitions from one isolated region of importance to another grow rare on a practical scale. Unfortunately, such rare-event sampling difficulties are themselves not rare events. They arise frequently, for example, in the treatment of activated processes at low temperatures.

[0038] The techniques disclosed herein for increasing the efficiency of simulations include two fundamental techniques: (1) "infinite swapping" of information within a limited set and (2) a "dual-chain" technique in which information is exchanged between different groupings. Each of these techniques is explained further below.

[0039] FIG. 17 is a schematic diagram that illustrates both the "infinite swapping" and "dual-chain" techniques. Figure 17 shows a first set of groups 1710 and a second set of groups 1720. Each of the groups extends from a lowest value, V_{lowest}, to a highest value, $V_{highest}$. The first set of groups 1710 includes N groups designated Group 1, Group 2, ... Group N. The second set of groups 1720 also includes N groups designated Group 1' (i.e., "Group 1 prime"), Group 2', ... Group N'. As shown, all the groups in the first set 1710 are distinct from all groups in the second set.

[0040] As an example, the groups can extend over a temperature range such that V_{lowest} represents 0.050 degrees and $V_{highest}$ represents 0.330 degrees, and such that 45 temperatures are selected to cover the intervening range from (0.050-0.210) in temperature steps of 0.005, and from (0.210-0.330) in steps of 0.010. The temperature values included in the groups for this example is summarized below in Table A.

[0041] **Table A**

First set of groups 1710	
Group identifier	*Temperatures included in the group*
Group 1	0.050, 0.055, 0.060
Group 2	0.065, 0.070, 0.075, 0.080, 0.085, 0.090, 0.095
Group 3	0.100, 0.105, 0.110, 0.115, 0.120, 0.125, 0.130
Group 4	0.135, 0.140, 0.145, 0.150, 0.155, 0.160, 0.165
Group 5	0.170, 0.175, 0.180, 0.185, 0.190, 0.195, 0.200
Group 6	0.205, 0.210, 0.220, 0.230, 0.240, 0.250, 0.260
Group 7	0.270, 0.280, 0.290, 0.300, 0.310, 0.320, 0.330
Second set of groups 1720	
Group identifier	*Temperatures included in the group*
Group 1'	0.050, 0.055, 0.060, 0.065, 0.070, 0.075, 0.080
Group 2'	0.085, 0.090, 0.095, 0.100, 0.105, 0.110, 0.115
Group 3'	0.120, 0.125, 0.130, 0.135, 0.140, 0.145, 0.150
Group 4'	0.155, 0.160, 0.165, 0.170, 0.175, 0.180, 0.185
Group 5'	0.190, 0.195, 0.200, 0.205, 0.210, 0.220, 0.230
Group 6'	0.240, 0.250, 0.260, 0.270, 0.280, 0.290, 0.300
Group 7'	0.310, 0.320, 0.330

[0042] In this example, the groups are characterized by the following properties:

- each set of groups includes each member of the discrete set of temperatures consisting of: {0.050, 0.055, 0.060, 0.065, 0.070, 0.075, 0.080, 0.085, 0.090, 0.095, 0.100, 0.105, 0.110, 0.115, 0.120, 0.125, 0.130, 0.135, 0.140, 0.145, 0.150, 0.155, 0.160, 0.165, 0.170, 0.175, 0.180, 0.185, 0.190, 0.195, 0.200, 0.205, 0.210, 0.220, 0.230, 0.240, 0.250, 0.260, 0.270, 0.280, 0.290, 0.300, 0.310, 0.320, 0.330}

- no group contains more than seven temperature values

- groups in the first set 1710 are distinct from groups in the second set 1720 (i.e., no group in the first set 1710 includes the identical set of temperatures included in any group of the second set 1710)

[0043] As explained below, the "infinite swapping" technique, which uses a symmetrized distribution and allows simultaneous exchange of information within a group of values, is impractical if carried out over too large a set of values. However, the "infinite swapping" technique is practical if carried out over a reasonable set of values. For example, seven or eight values is a reasonably sized set. Thus, in the example above, no group contains more than seven temperature values. As explained further below, when the technique illustrated in Figure 17 is performed, "infinite swapping" is allowed to occur within each individual group.

[0044] Although "infinite swapping" is practical when carried out over a reasonably-sized set of values, it is often desirable to simulate a physical system over a much larger set of values. In the example described above in connection with Figure 17, it is desired to simulate a physical system at forty-five different temperature values, i.e., 0.050, 0.055, ... 0.330 degrees. Forty-five different values is too large for "infinite swapping" to be practical.

[0045] As again explained further below, the "dual chain" technique provides a good approximation to "infinite swapping" extended over the full range of values of interest. As an example, the "dual chain" technique can be performed as follows:

[0046] • Infinite swapping analysis is performed within every group in the first set 1710 (i.e., infinite swapping analysis is performed for the values in Group 1, a separate infinite swapping

analysis is performed for the values in Group 2, ... and a separate infinite swapping analysis is performed for the values in Group N).

[0047] • Data generated in the infinite swapping analysis of the first set 1710 is then provided to the second set 1720 as illustrated by arrow 1730 (e.g., data generated for Groups 1 and 2 of the first set 1710 is used to seed a simulation of Group 1' in the second set 1720; data generated for Groups 2 and 3 of the first set 1710 is used to seed a simulation of Group 2' in the second set 1720; ... and data generated for Group N of the first set 1710 is used to seed a simulation of Group N' of the second set 1720.

[0048] • Infinite swapping analysis is then performed within every group in the second set 1720 (i.e., infinite swapping analysis is performed for the values in Group 1', a separate infinite swapping analysis is performed for the values in Group 2', ... and a separate infinite swapping analysis is performed for the values in Group N').

[0049] • Data generated in the infinite swapping analysis of the second set 1720 is then provided to the first set 1710 as illustrated by arrow 1740 (e.g., data generated for Group 1' of the second set 1720 is used to seed a simulation of Group 1 in the first set 1710; data generated for Groups 1' and 2' of the second set 1720 is used to seed a simulation of Group 2 in the first set 1710; ... and data generated for Groups (N−1)' and N' of the second set 1720 is used to seed a simulation of Group N of the first set 1710.

[0050] In this manner, infinite swapping analysis within one set of groups followed by exchange of information from a previous infinite swapping analysis to a new set of groups and infinite swapping analysis of that group, followed by another exchange of information is performed until the desired results of the simulation have been obtained.

[0051] The techniques of "infinite swapping" and "dual chain" have been described above in connection with simulating a physical system over a range of temperatures. However, the values V_{lowest} and $V_{highest}$ need not represent temperatures. As further examples, the simulations can be

performed over any parameter, or control variable, of interest such as pressure, density, volume, temperature, or others.

[0052] Also, in the example given above, each set 1710 and 1720 contains N groups and each group contains either three or seven values. In practice, each set need not contain the same number of groups and the groups need not contain the same numbers of values. Ideally though, no group will contain more values than would be practical for "infinite swapping" analysis within the group. Also, ideally, the groups in one set are distinct from the groups in another set such that the "dual chain" technique allows a group to receive information from a prior "infinite swapping" analysis that was performed over a range of values different from those contained within the group (e.g., as in the example above, Group 2' which contains values from 0.085-0.115 receives information from an infinite swapping analysis performed on groups 1 and 2, which covered the overlapping but distinct range of values 0.050...0.095).

[0053] The general technique described above includes: (a) infinite swapping analysis in every group of a first set of groups, followed by (b) using information from the first infinite swapping analysis to seed a simulation for groups in a second set of groups, followed by, (c) infinite swapping analysis in every group of the second set of groups, followed by (d) using information from the second infinite swapping analysis to seed a simulation for groups in the first set, followed by (e) a new infinite swapping analysis in every group of the first set. The process of infinite swapping analysis followed by exchange of information to a new set of groups can continue indefinitely or until desired results are obtained. It should be appreciated that the analysis need not alternate between two sets of groups as has been described above. As an alternative, data from the second infinite swapping analysis can be used to seed a simulation for a third set of groups, and so on.

[0054] As explained further below, the "infinite swapping" and "dual chain" techniques can be used broadly for the study of, e.g., nanoscale structures, catalysts, and geometry of proteins. For nanoscale structures, such as Lennard-Jones clusters, the use of grouping allows improved simulation of control values to be used when simulating each of the pairwise interactions of, e.g., a cluster of 10 gold atoms and 15 silver atoms. For catalysts, a similar simulation of the physical properties of each of the atoms in a reaction may be enabled, even for determining the frequency of extremely rare events, thereby facilitating the analysis of a proposed catalyst. For protein geometry, the electrical, nuclear and van der Waals interactions between a large number of atoms may be simulated using the improved performance of this technique, such that large-scale simulation of protein folding at different temperatures may be modeled.

[0055] FIG. 18 is a schematic diagram of a computer system, in accordance with some embodiments. Computing device 1800 contains processor 1801, memory 1802, grouping module 1803, swapping module 1805, and analysis module 1806. In one embodiment, data is received by device 1800 and grouped into a first grouping by grouping module 1803, and then handed off to swapping module 1805. Swapping module 1805 performs the logical control functions and transfers data between grouping module 1803 and analysis module 1806 to enable the "dual chain" technique. Analysis module 1806 performs one or more operations that may include symmetrization operations and that may include operations described below as infinite swapping or partial infinite swapping. Once analysis module 1806 is done processing the first grouping, it returns its information to swapping module 1805, which then may determine whether to perform additional groupings, perform additional analysis, or to terminate processing. In some embodiments, at least two groupings are made, thereby enabling dual-chain operation as described below.

[0056] FIG. 18 illustrates a single analysis module 1806. It will be appreciated that analysis module 1806 can be subdivided into separate analysis modules, such as a first analysis module (for performing infinite swapping analysis within a first set of groups) and a second analysis module (for performing infinite swapping analysis within a second set of groups). Alternatively, a common module may be used to perform infinite swapping analysis within multiple sets of groups.

[0057] To focus the present discussion, the problem is considered of estimating an average of the form

[0058]

$$\langle V \rangle = \frac{\int \mu(\mathbf{x}) V(\mathbf{x}) d\mathbf{x}}{\int \mu(\mathbf{x}) d\mathbf{x}}, \tag{1.1}$$

[0059] where x represents the coordinates of the system and V(x) is a property of interest such as the potential energy. As long as the probability distribution involved, $\mu(x)$, has a single "inherent structure', the sampling is straightforward and the associated numerical results are reliable. If, on the other hand, the probability distribution has multiple, isolated inherent structures, this ceases to be the case. In such cases <V> can be thought of as a sum of contributions over the various inherent structures,

[0060]

$$\langle V \rangle = \sum_{\alpha} \langle V \rangle_{\alpha} W_{\alpha} .$$

(1.2)

[0061] In Eq. (1.2) $\langle V \rangle_{\alpha}$ represents the average of V over the region corresponding to inherent structure α, $\mathfrak{R}_{\alpha}$

[0062]

$$\langle V \rangle_{\alpha} = \frac{\int_{\mathfrak{R}_{\alpha}} \mu(\mathbf{x}) V(\mathbf{x}) d\mathbf{x}}{\int_{\mathfrak{R}_{\alpha}} \mu(\mathbf{x}) d\mathbf{x}},$$

(1.3)

[0063] and W_{α} is the fraction of the statistical weight associated with that region:

[0064]

$$W_{\alpha} = \frac{\int_{\mathfrak{R}_{\alpha}} \mu(\mathbf{x}) d\mathbf{x}}{\int \mu(\mathbf{x}) d\mathbf{x}}.$$

(1.4)

[0065] Even though one does not typically decompose the required averages into explicit component form, as emphasized by Eq. (1.2) both intra and inter-inherent structure sampling are at issue in their calculation. The relevant point is that while conventional sampling methods do well for the former task, their performance for the latter can be unreliable.

[0066] Methods designed to deal with rare-event sampling issues have been developed and are briefly summarized in Section II. While many represent a significant improvement over basic methods, the need for more effective techniques remains an important area of current research. In the present work a new approach is described that is based on the use of

symmetrization. At first glance, the conscious introduction of this element into the discussion is counter-intuitive in that its presence is generally regarded as increasing rather than decreasing the complexity of the problem. As shown in the following sections, however, because it alters the inherent connectedness of the probability distributions involved, symmetrization represents a potentially useful tool in the treatment of rare-event problems.

[0067] The outline of the remainder of the paper is as follows: Section II presents both a brief overview of current approaches to the rare-event sampling problem and a heuristic discussion of the present technique. For reasons that will be discussed in greater detail in Section II, this method is denoted the "infinite swapping" or INS technique. The details of the INS approach and practical methods for its implementation are then outlined in Section III. Numerical applications that compare the performance of INS and parallel tempering methods for the calculation of thermodynamic properties of selected Lennard-Jones (LJ) clusters are discussed in Section IV. Finally, Section V summarizes the findings the instant application and indicates possible directions for future development.

[0068] Background

[0069] A number of methods for dealing with rare-event sampling problems have been developed and are discussed elsewhere. Common strategies include changes in trial moves, changes of measure and parallel tempering or replica exchange methods. Perhaps the simplest device for dealing with rare-event problems is to increase the length scale(s) associated with trial displacements. Although such increases generally result in diminished acceptance probabilities, they can nonetheless sometimes improve the ability of the associated random walks to overcome barriers. If one has sufficient physical insight concerning the nature of the relevant potential energy surfaces involved, "displacement vector Monte Carlo" techniques become an option and can be effective.

[0070] A general technique for dealing with rare-event issues is through a change of measure. For example, one can rewrite the average in Eq. (1.1) as

[0071]

$$\langle V \rangle = \frac{\left\langle \left(\frac{\mu(\mathbf{x})}{\tilde{\mu}(\mathbf{x})} \right) V(\mathbf{x}) \right\rangle_{\tilde{\mu}}}{\left\langle \left(\frac{\mu(\mathbf{x})}{\tilde{\mu}(\mathbf{x})} \right) \right\rangle_{\tilde{\mu}}}, \qquad (2.1)$$

[0072] where the averages on the right hand side of Eq. (2.1) are over an auxiliary probability distribution, $\tilde{\mu}(\mathbf{x})$, that can be chosen for sampling convenience. When adopting this approach one must guard against $\tilde{\mu}(\mathbf{x})$ choices that cause unacceptable increases in the associated variance. A number of strategies for $\tilde{\mu}(\mathbf{x})$ selection have been proposed.

[0073] A widely utilized method for dealing with rare-event sampling issues is the parallel tempering or replica exchange technique. This approach utilizes an expanded computational ensemble composed of systems corresponding to different values of one or more of the control variables. In the study of activated processes, for example, temperature is a natural parameter. Parallel tempering tries to overcome rare-event sampling issues by exchanging information between different portions of the simulation. Information from higher temperatures, where high-energy barriers are more easily crossed, is used to improve the sampling at lower temperatures, where such crossings are more difficult. Strategies for the selection of the tempering ensemble have been discussed.

[0074] Information transfer within the computational ensemble is essential to the parallel tempering approach. It is important to note that the nature of the exchange moves that are typically used to generate this transfer places practical limits on the rate of information flow achievable within the approach. For example, conventional, single-temperature coordinate displacements are typically augmented with trial moves that involve attempted swaps of system coordinates between temperatures. The percentage of swap attempts used is generally set using the Goldilocks - "not-too-small not-too-big" approach. If the percentage is too small, the information flow between temperatures is minimal and the resulting sampling benefits are negligible. If, on the other hand, the percentage is too large, the sampling again suffers. In the extreme limit, where all trial moves become attempted swaps, fixed system configurations are ineffectively toggled back and forth between the various temperature streams.

[0075] Using the theory of large deviations, it is possible to show that the performance of parallel tempering is, in principle, a monotonically increasing function of the swap rate. Empirical indications of such behavior have been noted. Were it possible to design such a procedure, this result suggests that the infinite swapping limit of parallel tempering would have desirable performance characteristics. This analysis shows that it is, in fact, possible to construct an alternative scheme, which coincides with the infinite swapping limit of parallel tempering in a distributional sense and which can be readily implemented. More generally, the method that emerges is a special case of a new class of rare-event sampling techniques.

[0076] The basic ideas that emerge from the large deviations analysis are shown by considering a simple example, a double well system whose potential energy is given by

$$V(x) = (x^2 - 1)^2. \tag{2.2}$$

[0077]

[0078] FIG. 1 is a schematic functional plot showing an exemplary potential energy function, in accordance with some embodiments. In Fig. (1) plots of the potential energy are shown, of the classical thermal distributions of the form

$$\pi(x,T) = e^{-V(x)/T}, \tag{2.3}$$

[0079]

[0080] for three temperatures, here chosen to be T = 0.05, 0.20,and 0.40. As expected, the normalized probability densities shown in Fig. (1) are increasingly "sparse" at lower temperatures where the scale of thermal fluctuations becomes small relative to the energy barrier that separates the wells.

[0081] FIG. 2 is a multidimensional plot showing isosurfaces, in accordance with some embodiments. In Fig. (2), the information of Fig. (1) is presented in a somewhat different manner. Specifically, a multi-variable distribution, μ_0, is defined as

[0082]

$$\mu_0(x,y,z,T_x,T_y,T_z) = \pi(x,T_x)\pi(y,T_y)\pi(z,T_z), \tag{2.4}$$

[0083] and assign the three temperatures used in Fig. (1) to particular coordinates, (Tx=0.05,Ty=0.20,Tz=0.40). The extent to which the various potential minima are linked is reflected in the structure of the isosurfaces of μ_0 shown in Fig. (2). The greatest bridging density is associated with the z-direction (highest temperature) and the least with the x-direction (lowest temperature).

[0084] FIG. 3 is a multidimensional plot showing symmetrized isosurfaces, in accordance with some embodiments. In Fig. (3) an isosurface plot is presented of a related distribution, μ_{xyz}, the fully symmetrized analog of the original. This (unnormalized) distribution is defined as

[0085]

$$\begin{aligned}\mu_{xyz}(x,y,z,T_x,T_y,T_z) = \; & \mu_0(x,y,z,T_x,T_y,T_z) + \mu_0(x,y,z,T_x,T_z,T_y) \\ & + \mu_0(x,y,z,T_y,T_x,T_z) + \mu_0(x,y,z,T_y,T_z,T_x). \\ & + \mu_0(x,y,z,T_z,T_x,T_y) + \mu_0(x,y,z,T_z,T_y,T_x)\end{aligned} \tag{2.5}$$

[0086] Because symmetrization removes the explicit linkage between temperature and coordinate labels, μ_{xyz} is more highly connected than μ_0. Thus, while both distributions contain related thermodynamic information, Figs. (2) and (3) suggest that the symmetrized version offers fewer rare-event sampling issues than the original. Whether or not this idea can be exploited and the factorial-scale growth of the computational complexity associated with total symmetrization can be avoided remains to be demonstrated.

[0087] FIG. 4a is a multidimensional plot showing partially-symmetrized isosurfaces, in accordance with some embodiments. FIG. 4b is a second multidimensional plot showing partially-symmetrized isosurfaces, in accordance with some embodiments. One possible strategy is suggested in Fig. (4). These plots display isosurfaces of two, partially symmetrized distributions, μ_{xy} and μ_{yz} defined for the present double-well example as

[0088]

$$\mu_{xy}(x,y,z) = \mu_0(x,y,z,0.20,0.40,0.05) + \mu_0(x,y,z,0.40,0.20,0.05) \tag{2.6}$$

[0089] and

[0090]

$$\mu_{yz}(x,y,z_z) = \mu_0(x,y,z,0.05,0.20,0.40) + \mu_0(x,y,z,0.05,0.40,0.20). \tag{2.7}$$

[0091] Although not as great as that seen for full symmetrization, Fig. (4) shows that partial symmetrization produces an increase in the connectedness of the associated distribution, a suggestion that the approach may offer a practical means of achieving the sampling benefits that are the goal of this disclosure.

[0092] Before discussing the details of the approach, it is useful to consider symmetrization from a somewhat different perspective. Equation (2.4) represents a typical parallel tempering ensemble; a simple product of Boltzmann terms each corresponding to a specified temperature. The mixing of coordinate and temperature information that underlies the method's improved sampling is generated indirectly, through the choice of trial moves. The structure of the symmetrized distribution defined in Eq. (2.5), on the other hand, intrinsically entangles temperature and coordinate information. All coordinate displacements, even those that would take place in a single temperature setting for the distribution in Eq. (2.4), thus occur in a Born-Oppenheimer like environment generated by the (infinitely rapid) multi-temperature swaps of information inherent in the symmetrization process.

[0093] Methods

[0094] In this section the INS approach is described and methods are developed for its practical implementation. The formal basis of the method is discussed elsewhere." An approach is presented in the context of a representative equilibrium problem, that of estimating the average potential energy of a system characterized by a set of coordinates, x, a known potential energy, V(x), and a specified probability distribution, $\pi(x, T)$. The objective is to compute the average potential energy at a specified temperature (or temperatures), T_k,

$$< V >_k = \frac{\int \pi(\mathbf{x},T_k)V(\mathbf{x})d\mathbf{x}}{\int \pi(\mathbf{x},T_k)d\mathbf{x}}. \tag{3.1}$$

[0095]

[0096] The joint probability distribution for N independent systems with coordinates, $\{x_n\}$ and a set of temperatures, $\{T_n\}$, n= 1, N (assumed to include T_k) is given by the product $\pi(x_1, T_1)\ \pi(x_2,T_2)...\ \pi(x_N,T_N)$ and its associated partition function is given by

$$Z = \prod_{n=1}^{N} \int d\mathbf{x}_n \pi(\mathbf{x}_n, T_n). \tag{3.2}$$

[0097]

[0098] The set of permutations of the temperatures between the coordinate sets is denoted by $\{P_n[\pi(x1,T1)\pi(x2,T2)...\pi(xN,TN)]\}$, where n=1,N! and the coordinate set label corresponding to the temperature T_k in the nth permutation by the function indx(n,k). In terms of these quantities, $\langle V\rangle_k$ can be written as the average of N! formally identical terms

[0099]

$$\langle V \rangle_k = \frac{1}{N!} \sum_n \frac{\int P_n[\pi(\mathbf{x}_1,T_1)\pi(\mathbf{x}_2,T_2)..\pi(\mathbf{x}_N,T_N)]V(\mathbf{x}_{indx(n,k)})d\mathbf{X}}{\int P_n[\pi(\mathbf{x}_1,T_1)\pi(\mathbf{x}_2,T_2)..\pi(\mathbf{x}_N,T_N)]d\mathbf{X}}, \tag{3.3}$$

[0100] where X represents the N sets of system coordinates, (Xj,X2, ...,XN). Because all of the normalization integrals for the terms in Eq. (3.3) are identical, the average involved can also be written as

[0101]

$$\langle V \rangle_k = \frac{\int \mu(\mathbf{X}) \left[\sum_n \rho_n(\mathbf{X}) V(\mathbf{x}_{indx(n,k)}) \right] d\mathbf{X}}{\int \mu(\mathbf{X}) d\mathbf{X}}, \tag{3.4}$$

[0102] where

[0103]

$$\rho_n(\mathbf{X}) = \frac{P_n[\pi(\mathbf{x}_1,T_1)\pi(\mathbf{x}_2,T_2)..\pi(\mathbf{x}_N,T_N)]}{\mu(\mathbf{X})}, \tag{3.5}$$

[0104] and where the (fully) thermally symmetrized distribution, μ(X), is given by

[0105]

$$\mu(\mathbf{X}) = \sum_n P_n[\pi(\mathbf{x}_1,T_1)\pi(\mathbf{x}_2,T_2)..\pi(\mathbf{x}_N,T_N)]. \tag{3.6}$$

[0106] The sum of the statistical weights defined in Eq. (3.5) is, by construction, equal to unity. For a specified temperature, the averages defined in Eqs. (3.1) and (3.4) are formally identical. However, as discussed in Section II, the increased connectedness of μ(x) suggests that μ(x) is a useful importance function and that Eq. (3.4) offers a potential computational advantage over Eq. (3.1) with respect to rare-event issues.

[0107] Equations (3.4)-(3.6) represent the full, N-temperature infinite swapping method. The invariant distribution involved is the sum of N! terms arising from all possible temperature permutations. If the number of temperatures is small, it is possible to sample this distribution directly. Even though complete INS sampling will become impractical beyond a modest number of temperatures, its consideration is useful since it provides both insights and benchmark examples that will prove useful for the development and validation of more generally applicable "partial" INS (PINS) approaches.

[0108] The fully symmetrized distribution is well defined and can be sampled using a variety of Monte Carlo techniques. One approach is shown in Appendix A and its application illustrated with a simple example. This approach may be valuable since, with suitable extension, it is well suited for both full and partial INS applications.

[0109] Simple card games played an important role in the discovery of modern Monte Carlo methods. They also provide a useful framework for understanding the partial swapping approach described in Appendix B. Generating all possible permutations of the symmetrized distribution

μ(x) is analogous to the problem of achieving all possible orderings in a deck of cards. Card shuffling techniques can be designed in a variety of ways. The most straightforward is to shuffle the entire deck at once. Alternatively, one can perform a multi-player shuffle in which players individually partition the deck into subsets, shuffle the cards within each of the subsets generated by their partitioning and then pass the deck to another player. As long as the partitionings involved do not share a common boundary, the resulting process will produce a complete shuffling. A simple, two-player, three-card example serves to illustrate the basic idea. Assume that when they have control of the deck, the players' actions are as follows:

[0110] • Player-1: leave top card in place, shuffle cards two and three

[0111] • Player-2: shuffle top two cards, leave card three in place.

[0112] Although neither player's individual actions would do so, the actions of the two players combined will shuffle the entire deck. Randomization is produced locally by the individual players and globally by their interaction. It is important to note that, even when the number of cards (temperatures) involved is large, the explicit randomization steps involve subsets whose sizes can be made controllably small. A general "dual" (two-player) sampling procedure is outlined based on this idea in Appendix B.

[0113] It is useful to note that the methods described in Appendix A and B contain internal diagnostics that can be helpful in numerical applications. Specifically, the statistical weights associated with the various permutations, Pn(x), contain information that provides insight into the nature of the sampling. If only a small number of the permutations in the relevant tempering ensemble have significant statistical weight, the sampling benefits associated with symmetrization will be minimal. This situation might arise, for example, if the spacings between the temperatures in the ensemble are too large. If, on the other hand, the ensemble temperatures are well chosen, then the statistical weight will be spread more broadly across the computational ensemble. The Shannon entropy associated with the ρ-values involved,

$$S_\rho = -\sum_n \rho_n \log(\rho_n),$$

[0114] (3,7)

[0115] is a measure of the dispersion of such ?-values and thus provides a useful device for monitoring the sampling.

[0116] Numerical Applications

[0117] A series of studies is presented below that are designed to explore the utility of the infinite and partial infinite-swapping approaches. Because they provide a common environment for few and many-body systems, exhibit a range of computational challenges, and have an extensive published literature, Lennard-Jones clusters have been chosen as the subject for these investigations. A series of relatively simple applications is used that illustrates the basic elements of the present approach and that provides a convenient means for comparing the performance of infinite swapping methods to that of current techniques. After these initial tests, a series of applications to systems of increasingly demanding rare-event character is presented in order to provide a basis for gauging the performance of the new techniques.

[0118] Unless otherwise stated, the following computational details apply to all studies discussed in the present section. Interaction potentials are assumed to be a sum of pairwise Lennard-Jones interactions of the form,

[0119]

$$v(r) = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right), \tag{4.1}$$

[0120] where □ and □ are energy and length scale parameters, respectively and r is the interparticle separation. Dimensionless energy and length scales, set by □ and □, are used throughout. Temperatures quoted are in dimensionless form ($k_B T/\varepsilon$) and all times are expressed in natural Lennard-Jones units, $(\square/m\sigma^2)^{1/2}$, where m is the atomic mass. Cluster evaporation is controlled by means of the customary addition of a confining potential to the pairwise Lennard-Jones interactions. The form of the confining potential used in the present study is

[0121]

$$V_c(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) = \sum_{i=1}^{N} v_c(\mathbf{r}_i), \tag{4.2}$$

[0122] where

$$v_c(\mathbf{r}) = \varepsilon\left(\frac{|\mathbf{r}-\mathbf{r}_{cm}|}{R_c}\right)^{20}, \qquad (4.3)$$

[0123]

[0124] and where r_{cm} and R_c are the cluster center of mass and constraining radius, respectively. In order to provide a uniform basis for comparison of the various sampling methods being discussed, a common technique, the smart Monte Carlo (SMC) method, is used to perform the necessary sampling moves. The choice of the SMC approach is based on a variety of considerations. At the practical level, SMC methods make use of familiar and readily available molecular dynamics techniques, thus making them relatively easy to implement. In addition, dynamically based SMC methods are well-suited for treating the correlated, many-particle barrier dynamics involved in these activated processes. In the SMC approach "moves" are generated by assigning Gaussian random moments, a characteristic of the temperature in question to all Cartesian degrees of freedom for the cluster and then propagating the classical equations of motion forward for a specified length of time via molecular dynamics methods. As defined, SMC "moves" thus involve the displacement of all cluster coordinates. Subsequent moves are generated by repeating this process with the last configuration serving as the starting point for the next move. Provided that the total system energy is conserved for the molecular dynamics segments, the configurations so produced provide a proper sampling of the Boltzmann distribution of positions at the specified temperature. Because the time scale for the molecular-dynamics segments involved is relatively short (on the order of the cluster vibrations), it is relatively easy to achieve an acceptable level of energy conservation. Fourth-order Runge-Kutta methods with step sizes of between 0.005 and 0.010 Lennard-Jones time units are utilized for all molecular dynamics integration tasks in the present work.

[0125] Table 1

k_BT/ε	INS	1-4/4-1	2-3/3-2	1-2-2/2-2-1	Metropolis
0.0500	-5.8370 (3)	-5.8374 (3)	-5.8375 (3)	-5.8378 (4)	-5.8366 (7)
0.0600	-5.8007 (3)	-5.8012 (4)	-5.8010 (4)	-5.8013 (4)	-5.8017 (9)
0.0800	-5.7219 (5)	-5.7223 (5)	-5.7227 (6)	-5.7225 (6)	-5.7225 (12)
0.1000	-5.6300 (8)	-5.6305 (10)	-5.6310 (9)	-5.6297 (10)	-5.6319 (17)
0.1200	-5.5191 (18)	-5.5184 (20)	-5.5184 (18)	-5.5190 (20)	-5.5178 (24)

[0126] Table 1 shows a series of average potential energies computed for the simple, four-atom LJ cluster at five temperatures. These results are computed for the various temperatures using a number of different approaches, including the ordinary Metropolis technique, INS, and three different PINS approaches. The numbers quoted in parentheses are the standard deviation estimates for the last digit quoted in the associated results. To facilitate comparison, all results are generated using 2 x 105 SMC moves of 0.50 time duration and a confining potential parameter of $R_c = 2.5$. Of these moves, S x 10^4 are used as warm-up. Because there is only one stable potential minimum, rare-event sampling issues are not a concern for this system. What is relevant is that in Table I that all methods yield consistent estimates of the average potential energy at the various temperatures involved. In particular, the average potential energies produced by the full INS method, various types of PINS sampling, and single-temperature Metropolis sampling all agree.

[0127] A comparison is now made of the performance of a number of different sampling methods for a series of increasingly demanding applications. FIG. 5a is a temperature profile plot for four relaxation simulations, in accordance with some embodiments. FIG. 5b is a relaxation curve plot using infinite swapping, in accordance with some embodiments. "Relaxation experiments," of the type illustrated in Fig. (5) for a simple 4-temperature 13-atom LJ cluster, are a convenient device for making such comparisons. The idea behind such numerical experiments is to impose a periodic temperature variation in the computational ensemble and to monitor the response of a calculated average property, in this case the average potential energy. One period of the temperature variations used in the present study is shown in Fig. (5a). As this cycle is repeated, potential energies corresponding to various numbers of SMC moves are accumulated and "relaxation" curves of the type shown in Fig. (5b) are generated. The results in Fig. (5b) are generated using 2000 heating/cooling cycles of the type displayed in Fig. (5a) using complete INS sampling. SMC moves of 0.50 duration and a confining potential parameter of $R_c = 2.5$ are used. For comparison purposes, the values of the average potential energies for the high and low-temperatures shown in Fig. (5b) are computed in separate INS simulations using 6 x 10^5 SMC moves. The rates at which the heating/cooling curves approach their limiting equilibrium values can be seen in Fig. (5b).

[0128] FIG. 6 is a cooling portion of the lowest-temperature curve in FIG. 5b, in accordance with some embodiments. Relaxation studies allow us to compare the rates of equilibration for different sampling methods. In Fig. (6), for example, the cooling portions of the lowest of the four temperatures shown in Fig. (5) are obtained for comparison using various sampling

methods, including INS sampling, parallel tempering, and conventional single-temperature Metropolis methods. The parallel tempering results shown in Fig. (6) are labeled with the percentages of trial swap moves involved. Specifically, the percentage stated corresponds to the probability that during each update of the coordinates there is an attempted parallel tempering exchange between one, randomly chosen, nearest-neighbor pair of temperatures. In this notation, conventional, single-temperature Metropolis results correspond to the zero swapping limit of parallel tempering. The other computational parameters used for the studies of Fig. (6) are the same as those in Fig. (5). To provide a common basis for comparison, all studies in Fig. (6), including the conventional Metropolis approach, utilize SMC sampling moves of duration 0.50. The observed rates of equilibration for the sampling methods vary significantly, with conventional, single-temperature Metropolis sampling being the slowest and the INS approach the fastest. In Fig. (6) that over the range typically utilized in parallel tempering simulations the rate of equilibration is an increasing function of the percentage of pair swap attempts.

[0129] FIG. 7 is a plot of a representative value on the parallel tempering relaxation curves of FIG. 6, in accordance with some embodiments. As discussed in Section II and as illustrated in Fig. (7), the improvement in the convergence rate (as measured by the value of <V> at a particular point on the relaxation curve) ultimately saturates and reverses as the percentage of attempted pair swaps is increased. The "optimal" swap percentage for this example is larger than that commonly utilized in tempering applications. It should be noted that "optimal" as described herein may refer to a relative degree of optimality with regards to the prior art, or to a degree of optimality within the framework of the disclosed invention, or to another meaning of the term.

[0130] Finally, an examination of the quality of the partial infinite swapping approach is shown below. FIG. 8 is a plot of a representative value on the parallel tempering relaxation curves of FIG. 6 using partial infinite swapping, in accordance with some embodiments. In Fig. (8) the cooling curves for the INS and parallel tempering results of Fig. (6) are compared with that obtained using a dual-chain PINS approach. Using the notation of Appendix B, both PINS chains are composed of two temperature blocks, a lowest temperature and a group of three higher temperatures for one chain and group of three lower temperatures and a single highest temperature for the other. The performance of this PINS(1-3/3-1) approach is quite similar to that of the full infinite swapping method. For example, the value of <V> for the PINS(I-3/3-1) cooling curve at $N_{move} = 100$, the point used in Fig. (7) to gauge the rate of parallel tempering equilibration, is -42.64, quite close to that produced by the full INS simulation (-42.66) and better than the "optimal" parallel tempering result of Fig. (7) (-42.49). Although encouraging,

conclusions concerning the performance and general utility of the PINS method await more serious tests.

[0131] The 38-atom Lennard-Jones cluster is an appreciably more complex system than those considered thus far. The potential energy surface for this system exhibits a multi-funnel structure in which the global and lowest-lying local minimum differ in energy by an amount that is small relative to the barrier that separates them. This and the presence of roughly 4×10^4 local minima are indications that the LJ-38 system presents non-trivial computational challenges.

[0132] Although other, more effective methods are shown later, the LJ-38 investigations are first shown with a series of applications that utilize a rather basic version of the PINS approach, the "floating" partial swapping method. As discussed in Appendix B, this simple approach utilizes a symmetrized band of temperatures that moves randomly within the larger computational ensemble. Using these methods, the issue of equilibration in the LJ-38 system is examined.

[0133] FIG. 9 is a loop average potential energy plot for a 38-atom Lennard-Jones cluster, in accordance with some embodiments. In Fig. (9) loop averages (100 SMC moves per loop) are compared for two different six-temperature floating PINS simulations. Both simulations use ensembles that contain 45-temperatures that span the range from (0.050-0.210) in steps of 0.005 and the range from (0.210 - 0.330) in steps of 0.010. SMC moves of 1.00 duration and a confining potential parameter of $R_c = 2.65$ are used. Loop averages shown in Fig. (9) correspond to the potential energies for 18 of the 45 temperatures and cover the range from 0.050-0.180. Results shown on the left in Fig. (9) are from a simulation that is initiated using a randomly chosen, high-temperature "melt" configuration (a down-anneal) while those on the right are from a second simulation that is initiated using the known global minimum configuration (an up-anneal). For ease of visual comparison of the final results, the output for the up-anneal simulation is displayed in reverse in Fig. (9). That is, the results of the two simulations begin at the edges and end in the center of the figure. The loop averages generated in the up and down anneals for the various temperatures are consistent (i.e. merge as they approach the center of the figure).

[0134] FIG. 10 is a heat capacity plot for a 38-atom Lennard-Jones cluster, in accordance with some embodiments. As further evidence of this consistency, the heat capacities computed using the last 20,000 loops or 2×10^6 SMC moves of both simulations are compared in Fig. (10). The heat capacities for these two simulations, including the "shoulder" region between temperatures of 0.10 and 0.15, are in good agreement.

[0135] FIG. 11 is a loop average potential energy plot for a 38-atom Lennard-Jones cluster with an alternate simulation start point, in accordance with some embodiments. FIG. 12 is a heat capacity plot for a 38-atom Lennard-Jones cluster with an alternate simulation start point, in accordance with some embodiments. In Figs. (11) and (12) additional LJ-38 equilibration and heat capacity studies are presented. In these figures the up-anneal results of Figs. (9) and (10) are replaced with analogous six-temperature floating PINS results in which the initial configuration is taken to be that of the lowest-lying LJ-38 icosahedral isomer instead of the global minimum. Significant equilibration difficulties have been reported for parallel tempering simulations involving this particular isomer. The quality of the PINS results in Figs. (11) and (12) are similar to those shown in Figs. (9) and (10), a further indication that PINS methods are effective in coping with the rare-event sampling issues presented by the LJ-38 system.

[0136] The LJ-38 results shown in Figs. (9-12) indicate that the floating PINS approach is a useful technique. It is important to note, however, that other, more effective methods are available. As an illustration, Fig. (13) shows the results of a series of numerical relaxation experiments for the LJ-38 system. These are obtained using parallel tempering (16% swap rate), floating PINS and dual-chain PINS approaches for the 45 temperature ensemble described previously. For simplicity, only results for the lowest temperature in the computational ensemble are shown. Because the floating PINS simulation is based on a six-temperature symmetrized block, a dual-chain PINS study is shown in which sampling chains are composed of the same size blocks. Specifically, each of the chains is composed of seven, six-temperature blocks plus one smaller block of three temperatures for a total of 45 temperatures. In one chain, the three-temperature block is made up of the lowest three temperatures, in the other chain the highest three. Heating and cooling profiles used are of the generic type illustrated in Fig. (5a). In the present study, however, the cycles consist of 1200 SMC moves, each of one unit Lennard-Jones time duration. The cooling segment is taken as the portion of the cycle from moves 200 to 800 with the remainder being the heating portion. During the cooling portion of the cycle the 45 temperatures in the ensemble cover the range from (0.050-0.210) in temperature steps of 0.005, and from (0.210 - 0.330) in steps of 0.010 while during the heating portion of the cycle temperatures less than or equal to 0.150 are set equal to 0.150. It should be noted that for a fixed block size, the computational effort in the dual-chain PINS approach grows linearly in the total number of temperatures. All simulations utilize a constraining radius of $R_c = 2.65$ and are obtained using 600 thermal cycles.

[0137] FIG. 13 is a relaxation plot for a 38-atom Lennard-Jones cluster, in accordance with some embodiments. In Fig. (13), equilibration is achieved by both the floating and dual-chain PINS approaches and that the rates for both are significantly better than those of parallel tempering. In fact, evidence of incomplete equilibration is seen in the parallel tempering results of Fig. (13). The rates of equilibration for the dual-chain PINS approach are also significantly better than those for the simple floating PINS approach. It is important to note that this improvement in equilibration rates is achieved with a relatively modest increase in computational effort. For example, the computing times required for the floating and dual-chain PINS results in Fig. (13) are only 9% and 12% more than those required for the corresponding parallel tempering results, respectively.

[0138] Finally, a 31-atom Lennard-Jones cluster is considered, a system that offers a number of interesting features. As discussed by Doye, et al., the potential energy surface for this system is relatively complex with many, nearly degenerate local minima separated by large barriers. This complex energy landscape makes locating the global minimum structure difficult and produces a low-temperature heat capacity feature that represents a challenging computational objective.

[0139] FIG. 14 is an equilibration plot for an LJ-31 system, in accordance with some embodiments. In Fig. (14) down and up-anneal studies are presented for the LJ-31 system. The results shown are obtained using two PINS(4/54) simulations. As with the previous LJ-38 studies, the down-anneal starts from a randomly chosen, high-temperature "melt" configuration while the up-anneal is initiated using the known global minimum. The 54 temperatures in the ensemble are distributed geometrically in the range (0.01, 0.35). Loop averages for 18 temperatures over the range (0.01,0.0535) are shown in Fig. (14). These simulations contain various numbers of loops that consist of 100 SMC moves per loop, with each SMC move being of unit time duration. The confining radius used in these LJ-31 studies is $R_c = 3.00$, the same as that in Ref. (33). The down-anneal simulation uses 37,000 total loops with the final 25,000 being used for data collection. Because the warm-up period involved is appreciably smaller, the up-anneal uses 27,000 total loops, again with the final 25,000 being for data collection.

[0140] FIG. 15 is a heat capacity plot for an LJ-31 system, in accordance with some embodiments. The heat capacities for the LJ-31 system computed from these two simulations are shown in Fig. (15). As was the case with the analogous LJ-38 studies, the agreement of the heat capacities produced by the up and down-anneal simulations is quite good, including the region of the low-temperature heat capacity peak near T = 0.027. The potential energy

fluctuations associated with rare-event hops between the various local minima that produce this low-temperature heat capacity feature are evident in Fig. (14) in the potential energy range of roughly -132.5 ± 0.5.

[0141] Discussion and Summary

[0142] The rare-event sampling problem is a troublesome and ubiquitous one. The present work develops an approach for dealing with such problems that is based on the use of symmetrization. In addition, the results have been shown for an initial series of applications to a representative class of problems, the calculation of the thermodynamic properties of Lennard-Jones clusters. Preliminary findings suggest that the infinite swapping approach is an effective tool for dealing with rare-event problems and represents an improvement over parallel tempering.

[0143] It is important to note that, while both the infinite swapping and parallel tempering approaches utilize an expanded computational ensemble to improve their rare-event sampling performance, there is an important distinction between the two approaches. In parallel tempering the invariant distribution is a simple product of individual Boltzmann components. The transfer of temperature-related information used to improve the rare event sampling is achieved by expanding the space of possible trial moves to include swaps of data between the various temperature streams. In contrast, the invariant distribution for the infinite swapping approach is a thermally symmetrized sum of Boltzmann-like terms. This distribution is more highly connected than the original, unsymmetrized form. Moreover, its basic structure intrinsically entangles temperature and coordinate information thereby increasing the flow of temperature-related information within the computational ensemble. This increased flow of information is at the core of the infinite swapping approach. Practical methods for sampling the associated distributions have been presented and discussed.

[0144] Beyond the classical statistical-mechanical applications considered in the present work, infinite swapping methods would appear to have other areas of potential utility. One obvious area is that of minimization. Preliminary applications suggest that the increased flow of temperature-related information makes INS-inspired approaches useful for such applications and that further investigations are warranted. Quantum simulations, where the problems of sampling symmetrized and sparse distributions arise naturally in path-integral simulations, are other areas of possible payoff.

[0145] Appendix A

[0146] An approach called the infinite swapping sampling approach is described by illustrating its application to the calculation of a representative thermodynamic property, the average potential energy at a specified temperature, T_k. For simplicity, this temperature is one of the set of ensemble temperatures $\{T_j\}$, j=1,N. The *N* sets of system coordinates, $(x_1,x_2, ...,x_N)$, are collectively denoted as X. The fully symmetrized distribution is defined in Eq. (3.6) and the statistical weight, $\rho_n(x_1,x_2, ...,x_N)$ for each of the *N*! possible permutations is given by Eq. (3.5). Assuming that the configuration at the *m*th step is denoted by $(x_{1,m},x_{2,m}, ...,x_{N,m})$, or as simply X_m, the sampling process includes the following steps:

[0147] • calculate the ρ-weights for the current configuration, $\{\rho_n(X_m)\}$, for all permutations in the N-temperature ensemble, n=1,N!;

[0148] • select one of the ρ-weights randomly according to its size;

[0149] • generate a new configuration, X_{m+1}, by making single-temperature moves for each of the coordinate sets using the temperature-coordinate associations inherent in the randomly chosen permutation;

[0150] • calculate the ρ-weights for all permutations in the N-temperature ensemble for the new configuration, $\{\rho_n(X_{m+1})\}$,n=1,N!;

[0151] • update the accumulating sum being used to assemble the <V>k average by adding $\Delta_{m+1}(k) = \sum_{j=1}^{N} V(x_{j,m+1})\Gamma(j,k)$ to the kth sum, where Γ(j,k) is equal to the sum of the ρ-weights for all permutations in which coordinate set j is associated with temperature Tk; and

[0152] • repeat the process using the new configuration, X_{m+1}, as input.

[0153] As an illustration, the above process is followed through a hypothetical cycle for a one-dimensional system and a three-temperature ensemble. If the permutation in which x_K is associated with temperature 1, x_L with temperature 2, etc. is denoted by [K,L,...], then the six possible permutations arising from the three-temperature ensemble are given by

[0154] n Pn

[0155] 1 [1,2,3]

[0156] 2 [1,3,2]

[0157] 3 [2,1,3]

[0158] 4 [2,3,1]

[0159] 5 [3,1,2]

[0160] 6 [3,2,1]

[0161] Given a particular configuration, $(x_{1,m},x_{2,m},x_{3,m})$ the ρ-values associated with each of the possible permutations are first evaluated, and then one of these permutations is selected randomly in proportion to its size. For purposes of illustration, permutation 3 in the above table is the permutation selected. Single-temperature Monte Carlo moves would then be made in which coordinate 2 would be advanced using temperature T1, coordinate set 1 using temperature T2, and coordinate 3 using temperature T3. The potential energies, $\{V(x_j)\}$, and ρ-weights, $\{\rho_n(x_j)\}$, j=1,3, n=1,6 corresponding to the new coordinates would be computed and the three accumulating sums being used in the calculation of the potential energy averages would be updated as

[0162]

$$\begin{aligned}
\mathrm{Sum}(T_1) &\rightarrow \mathrm{Sum}(T_1) + V(x_1)(\rho_1 + \rho_2) + V(x_2)(\rho_3 + \rho_4) + V(x_3)(\rho_5 + \rho_6) \\
\mathrm{Sum}(T_2) &\rightarrow \mathrm{Sum}(T_2) + V(x_1)(\rho_3 + \rho_5) + V(x_2)(\rho_1 + \rho_6) + V(x_3)(\rho_2 + \rho_4) \\
\mathrm{Sum}(T_3) &\rightarrow \mathrm{Sum}(T_3) + V(x_1)(\rho_4 + \rho_6) + V(x_2)(\rho_2 + \rho_5) + V(x_3)(\rho_1 + \rho_3).
\end{aligned}$$

[0163] Appendix B

[0164] In this appendix the partial INS sampling approach is described. As in Appendix A, an N-temperature ensemble is assumed and the objective is the calculation of thermodynamic properties such as the average potential energy. The methods outlined herein are designed for the calculation of such single-temperature properties. With suitable extension more general properties, for example those that depend on the joint distribution of multiple temperatures, can also be treated.

[0165] The procedure described herein is a "dual-chain" process in which the full, N-temperature set is partitioned into blocks in two distinct ways. Unlike the full INS approach, symmetrization is partial and is confined to the various temperature blocks that make up the two chains. The two chains are labeled α and β and denote the temperatures for block-1 of chain-α as $\{T_j^{\alpha_1}\}$, those for block-2 of the chain-β as $\{T_j^{\beta_2}\}$, etc. The system configuration at the *m*th step in the sampling is denoted as $(x_{1,m}, x_{2,m}, \ldots, x_{N,m})$ or as simply X_m. When necessary, partition and temperature block labels can also be included. For example, the coordinates at the mth sampling step of block-1 of chain-α are labeled as $\mathbf{X}_m^{\alpha_1}$, for block-2 of chain-β as $\mathbf{X}_m^{\beta_2}$, and so on.

Unless stated otherwise the partitionings of chains α and β are assumed to remain fixed during the simulation.

[0166] FIG. 16 is a schematic diagram of the partial infinite swapping approach, in accordance with some embodiments. The overall structure of the dual-chain PINS sampling method is depicted schematically in Fig. (16). Local mixing within blocks is produced by symmetrization while larger-scale mixing is generated by the exchange of information between chains. The individual chains are consistent with multiple invariant distributions, namely those that have the proper relative weights of the permutations within the various blocks. The unique distribution that the chains share is the fully symmetrized distribution. Consequently, while neither chain will do so individually, the two chains working in tandem will, if suitably designed, generate a proper sampling of the fully symmetrized distribution, Eq. (3.6). It is important to note that the PINS approach avoids dealing with all possible permutations simultaneously. Because the two chains utilize different partitionings, a proper sampling requires that rules be specified both to govern the construction of the required thermodynamic averages from the information produced by the individual chains and to control the "hand-off" of data between them. In these applications thermodynamic information emerges in the form of update increments that are accumulated to produce the average potential energies at the various temperatures. In Fig. (16) the potential energy increments for the temperatures in the first block of chain-α at step m+1 are denoted as $\Delta_{m+1}(k_{\alpha 1})$, those at step m+2 for the temperatures in the second block of chain-β as $\Delta_{m+1}(k_{\beta 1})$, and so on. The PINS sampling process is first described in general terms and then illustrate its use by considering a particular example. Assuming that the m+1st step starts with temperature block-1 of chain-a, the detailed steps in the "PINS cycle" depicted in Fig. (16) are as follows:

[0167] • define $Y_m = X_m^{\alpha_1}$;

[0168] • calculate the ρ-weights, $\{\rho_n(Y_m)\}$, n=l, $N(\alpha_l)$, for all $N(\alpha_l)$ permutations in block

[0169] • select one of the ρ-weights randomly in proportion to its size;

[0170] • generate a new configuration, Y_m+ 1, using the temperature-coordinate associations inherent in the randomly chosen permutation;

[0171] • calculate the new ρ-weights, $\{\rho_n(Y_{m+1})\}$, n=l, $N(\alpha_l)$, for all $N(\alpha_l)$ permutations in block α_l;

[0172] • calculate the potential energy update increments for the temperatures in the block $\Delta_{m+1}(k) = \sum_{\lambda=1}^{N(\alpha_l)} \Lambda(\lambda)^{w+1} L(\lambda, k)$, where r(j,k) is equal to the sum of the ρ-weights for all permutations of block α_l in which coordinate set j is associated with temperature T_k;

[0173] • select one of the newly generated ρ-weights randomly in proportion to its size; and

[0174] • use the temperature-coordinate associations in the randomly selected permutation to generate the coordinates for block α_l at the m+1st step, $X_{m+1}^{\alpha_l}$ to be handed off to the complementary chain.

[0175] These steps are then repeated for the remaining blocks in chain-α. After all chain-α steps are complete, the $X_{m+1}^{\alpha_l}$ coordinates for the various blocks are merged to form the new configuration, X_{m+1}, and this configuration is passed on to chain-β as indicated in Fig. (16). Chain-β then produces a new configuration, X_{m+2}, which is then passed back to chain-α, etc. and the process is repeated as necessary.

[0176] The PINS procedure is illustrated by considering its application to the three-temperature example of Appendix A using a dual-chain PINS(I-2/2-I) approach. This designation indicates that chain-α consists of a single-temperature block at temperature TI and a two-temperature, symmetrized block involving the temperatures T2 and T3. In chain-β, on the other hand, the two-temperature block is made up of temperatures TI and T2 while the single temperature block involves only T3. This system is the analog of the three-card example discussed near the end of Section III. For simplicity, the configuration at the mth sampling step is given by $Xm= (x_{1,m}, x_{2,m}, x_{3,m})$ and that the next sampling move involves chain-α. Defining Ym= Xm, the blocks and the p-values for the permutations in the two blocks in chain-α are evaluated numerically as:

[0177]

Block	n	$\rho_n(\mathbf{Y}_m)$
1	1	1
2	1	$\pi(y_{2,m},T_2)\pi(y_{3,m},T_3)/(\pi(y_{2,m},T_2)\pi(y_{3,m},T_3) + \pi(y_{3,m},T_2)\pi(y_{2,m},T_3))$
	2	$\pi(y_{3,m},T_2)\pi(y_{2,m},T_3)/(\pi(y_{2,m},T_2)\pi(y_{3,m},T_3) + \pi(y_{3,m},T_2)\pi(y_{2,m},T_3))$,

[0178] where n(y,T) is the single-temperature Boltzmann distribution defined by Eq. (2.3). In block-1 of chain-a, the coordinate-temperature association involves y_1 and T_1. Consequently,

$Y_{1,m+1}$ is generated from y_1 , m using the temperature T_1 and any suitable, single-temperature Monte Carlo procedure. Smart Monte Carlo methods are used for the applications in Section IV. Since there is only one coordinate and temperature for block-1, $x_{1,m+1} = y_{1,m+1}$ and the potential energy increment is $\Delta_{m+1}(T_1) = V(y_{1,m+1})$. To generate the sampling moves for block-2 of chain-α, evaluation of the block's ρ-values, ρ_1 and ρ_2, occurs for the variables $(y_{2,m}, y_{3,m})$ and one of the ρ-values is selected randomly in proportion to its size. For illustration purposes ρ_2 is assumed to be selected. Continuing, a Monte Carlo move is made from $y_{2,m}$ to $y_{2,m+1}$ and $y_{3,m}$ to $y_{3,m+1}$ using the temperature-coordinate associations in the selected term. From the last line of the table above, the temperature-coordinate associations in $\rho_2$is seen to involve the pairings y_2-T_3 and y_3-T_2. Consequently, y_2 is advanced at a temperature T_3 and y_3 at a temperature T_2. Had PI been chosen instead, the y_2 advance would have been made using the temperature T_2 and the corresponding y_3 advance using the temperature T_3. Using the new configuration $y_{m+1} = (y_{2,m+1}, y_{3,m+1})$, re-evaluation occurs of the block-2 ρ-values, ρ_1 (y_{m+1}) and $\rho_2(y_{m+1})$. These newly generated ρ-values are then used to produce the potential energy increments for T_2 and T_3 as well as the values of $(x_{2,m+1}, x_{3,m+1})$ that will be "handed-off" to chain-β. The potential energy increments are given by

[0179]

$$\Delta_{m+1}(T_2) = V(y_{2,m+1})\,\rho_1(Y_{m+1}) + V(y_{3,m+1})\,\rho_2(Y_{m+1})$$
$$\Delta_{m+1}(T_3) = V(y_{2,m+1})\,\rho_2(Y_{m+1}) + V(y_{3,m+1})\,\rho_1(Y_{m+1})$$

[0180] To generate the values of $(x_{2,m+1}, x_{3,m+1})$ that, along with $x_{1,m+1}$, will be passed on to chain-β, selection of one of the new ρ-values randomly in proportion to its size and assignment of the $(x_{2,m+1}, x_{3,m+1})$ values based on the coordinate-temperature associations in the selected term is performed. The results for the two possible selections are

[0181]

term selected	y-T associations	x-assignments
$\rho_1(Y_{m+1})$	y_2-T_2, y_3-T_3	$x_{2,m+1} = y_{2,m+1}$ $x_{3,m+1} = y_{3,m+1}$
$\rho_2(Y_{m+1})$	y_2-T_3, y_3-T_2	$x_{2,m+1} = y_{3,m+1}$ $x_{3,m+1} = y_{2,m+1}$

[0182] The final x-values so generated, $x_{m+1} = (x_{1,m+1},x_{2,m+1},x_{3,m+1})$, are then handed off as input to chain-β and an analogous series of steps are made using that chain's partitioning. The x_{m+2} values produced by the β-chain are then handed back into the α-chain and the entire dual process is repeated.

[0183] Partitionings that are composed of multiple blocks that contain the same, even number of temperatures in combination with a single block that contains half that number of temperatures have been used for convenience. To assure that they share no common temperature boundaries, the smaller block in one chain is made up of the lowest temperatures of the ensemble while in the other it involves the highest. The major block size and total number of temperatures thus serve as identifiers of the sampling chains used in the simulation. In this more compressed notation, for example, the previous PINS(I-2/2-1) illustration is designated as a PINS(2/3) simulation.

[0184] One has significant flexibility in designing many of the details of the PINS sampling approach. Decisions concerning the potential merits of many of these various options await further study. One option is the choice of the single-temperature sampling method. Although all numerical examples presented here use smart Monte Carlo methods.r" any technique that provides a proper sampling of the Boltzmann distributions involved can be utilized for the single-temperature sampling tasks. While the infinite sampling approach is guaranteed to perform better than parallel tempering, the degree of improvement can depend on the particular sampling used for the single-temperature moves. Another example is the possible use of multi-chain methods. It is straightforward to formulate multi-chain sampling methods that are generalizations of the dual-chain techniques utilized in the present studies. Whether or not these,

either by themselves or in combination with particular partitioning strategies, offer a computational advantage is an interesting and open question. Another example is the use of multi-step methods. In the present work only a single sampling step is made within each of the chains before handing-off the calculation to the complementary chain. It is important to note that multiple steps are permitted, a fact that may prove of practical significance in the context of parallel implementations of the PINS approach. Finally, the partitionings of the various chains need not remain fixed during the simulation. In the present work, for example, the "floating" PINS applications are implemented using a dynamical version of dual-chain sampling in which the chain partitions vary as a single, symmetrized block of temperatures moves randomly throughout the larger computational ensemble.

[0185] Figure Captions

[0186] Figure 1:

[0187] Plots of V(x) for the double-well example (Eq. (2.2)) and three normalized Boltzmann distributions (Eq. (2.3)) corresponding to T= 0.05, 0.20, 0.40.

[0188] Figure 2:

[0189] Plot of the isosurface $\mu_0(x,y,z) = 0.05$, where μ_0 is defined in Eq. (2.4).

[0190] Figure 3:

[0191] Plot of the isosurface $\mu_{xyz}(x,y,z) = 0.05$, where μ_{xyz} is defined in Eq. (2.5).

[0192] Figure 4:

[0193] (a) Plot of the isosurface $\mu_{xy}(x,y,z) = 0.05$, where μ_{xy} is defined in Eq. (2.6), (b) Plot of the isosurface $\mu_{yz}(x,y,z) = 0.05$, where μ_{yz} is defined in Eq. (2.7).

[0194] Figure 5:

[0195] (a) Temperature profiles for the four-temperature, LJ-13 relaxation simulations, (b) <V> values for the LJ-13 infinite swapping relaxation simulations performed using the temperature profiles in Fig. (5a).

[0196] Figure 6:

[0197] The cooling portion of the of the lowest of the four-temperature studies of Fig. (5b). Results correspond to infinite swapping (INS), parallel tempering (PT), and ordinary Metropolis sampling. The dashed line corresponds to the numerically exact value of<V> for T = 0.08.

[0198] Figure 7:

[0199] A plot of a representative value on the parallel tempering relaxation curves of Fig. (6) (Nmove = 100) as a function of the attempted pair swap probability (Pswap) .

[0200] Figure 8:

[0201] As in Fig. (6) with the addition of the partial swapping results (PINS).

[0202] Figure 9:

[0203] Loop averages (100 moves/loop) of the potential energy of a 38-atom LJ cluster. Results are obtained using a 6-temperature floating PINS simulation starting from a "melt" configuration (left edge to center) and from a second simulation starting from the known global minimum geometry (right edge to center). The bands correspond to the potential energies of 18 temperatures from the range (0.05, 0.18).

[0204] Figure 10:

[0205] Heat capacities for the LJ-38 system. The "down" ("up") labels refer to the corresponding simulations of Fig. (9).

[0206] Figure 11:

[0207] As in Fig. (9), except that the right hand portion of the plot corresponds to results obtained starting with the icosahedral isomer instead of the global minimum energy structure.

[0208] Figure 12:

[0209] Heat capacities for the LJ-38 system extracted from the last 20,000 loops of the simulations of Fig (11).

[0210] Figure 13:

[0211] Relaxation simulations for LJ-38. The <V> values shown are for the lowest temperature of the 45-temperature ensemble (see text for details) and are obtained using parallel tempering (PT), a "floating" sixtemperature partial swapping approach, and dual-chain PINS(6/45) methods. The numerically exact results for <V> = 0.05 are shown for companson.

[0212] Figure 14:

[0213] PINS(4/54) equilibration studies for the LJ-31 system. Temperatures in the ensemble are distributed geometrically in the range (0.01,0.35). Loop averages for 18 temperatures in the range (0.01,0.0535) are shown. Results on the left (right) side of the plot are initialized using a high-temperature melt configuration (the known global minimum geometry).

[0214] Figure 15:

[0215] The heat capacities of the LJ-31 system as a function of temperature extracted from the final 25,000 loops of the PINS(4/54) simulations of Fig. (14).

[0216] Figure 16:

[0217] A schematic depiction of the PINS sampling approach utilized in the present studies.

[0218] Appendix C

[0219] The disclosure below provides further mathematical details regarding the disclosure.

1. Introduction. The problem of computing integrals with respect to Gibbs measures occurs in chemistry, physics, statistics, engineering, and elsewhere. In many situations, there are no viable alternatives to methods based on Monte Carlo. Given an energy potential, there are standard methods for constructing a Markov process whose unique invariant distribution is the associated Gibbs measure, and an approximation is given by the occupation or empirical measure of the process over some finite time interval [10]. However, a weakness of these methods is that they may be slow to converge. This happens when the dynamics of the process do not allow all important parts of the state space to communicate easily with each other. In large scale applications this occurs frequently since the potential function often has complex structures involving multiple deep local minima.

An interesting method called "parallel tempering" has been designed to overcome some of the difficulties associated with rare transitions [4, 6, 21, 22]. In this technique, simulations are conducted in parallel at a series of temperatures. This method does not require detailed knowledge of or intricate constructions related to the energy surface and is a standard method for simulating complex systems. To illustrate the main idea, we first discuss the diffusion case with two temperatures. Discrete time models will be considered later in the paper, and there are obvious analogues for discrete state systems.

Suppose that the probability measure of interest is $\mu(dx) \propto e^{-V(x)/\tau_1}dx$, where τ_1 is the temperature and $V : \mathbb{R}^d \to \mathbb{R}$ is the potential function. The normalization constant of this distribution is typically unknown. Under suitable conditions on V, μ is the unique invariant distribution of the solution to the stochastic differential equation

$$dX = -\nabla V(X)dt + \sqrt{2\tau_1}dW,$$

where W is a d-dimensional standard Wiener process. A straightforward Monte Carlo approximation to μ is the empirical measure over a large time interval of length T, namely,

$$\frac{1}{T}\int_B^{T+B} \delta_{X(t)}(dx)dt,$$

where δ_x is the Dirac measure at x and $B > 0$ denotes a "burn-in" period. When V has multiple deep local minima and the temperature τ_1 is small, the diffusion X can be trapped within these deep local minima for a long time before moving out to other parts of the state space. This is the main cause for the inefficiency.

Now consider a second, larger temperature τ_2. If W_1 and W_2 are independent Wiener processes, then of course the empirical measure of the pair

$$dX_1 = -\nabla V(X_1)dt + \sqrt{2\tau_1}dW_1, \tag{1.1}$$
$$dX_2 = -\nabla V(X_2)dt + \sqrt{2\tau_2}dW_2$$

gives an approximation to the Gibbs measure with density

$$\pi(x_1, x_2) \propto e^{-\frac{V(x_1)}{\tau_1}} e^{-\frac{V(x_2)}{\tau_2}}. \tag{1.2}$$

The idea of parallel tempering is to allow "swaps" between the components X_1 and X_2. In other words, at random times the X_1 component is moved to the current location of the X_2 component, and vice versa. Swapping is done according to a state dependent intensity, and so the resulting process is actually a Markov jump diffusion. The form of the jump intensity can be selected so that the invariant distribution remains the same, and thus the empirical measure of X_1 can still be used to approximate μ. Specifically, the jump intensity or swapping intensity is of the Metropolis form $ag(X_1, X_2)$, where

$$g(x_1, x_2) = 1 \wedge \frac{\pi(x_2, x_1)}{\pi(x_1, x_2)} \tag{1.3}$$

and $a \in (0, \infty)$ is a constant. Note that the calculation of g does not require the knowledge of the normalization constant. A straightforward calculation shows that π is the stationary density for the resulting process for all values of a (see (2.2)). We refer to a as the "swap rate" and note that as a increases, the swaps become more frequent.

The intuition behind parallel tempering is that the higher temperature component, being driven by a Wiener process with greater volatility, will move more easily

between the different parts of the state space. This "ease of movement" is transferred to the lower temperature component via the swapping mechanism so that it is less likely to be trapped in the deep local minima of the energy potential. This, in turn, is expected to lead to more rapid convergence of the empirical measure to the invariant distribution of the low temperature component. There is an obvious extension to more than two temperatures.

Although this procedure is remarkably simple and needs little detailed information for implementation, relatively little is known regarding theoretical properties. A number of papers discuss the efficiency and optimal design of parallel tempering [8, 16, 15]. However, most of these discussions are based on heuristics and empirical evidence. In general, some care is required to construct schemes that are effective. For example, it can happen that for a given energy potential function and swapping rate, the probability for swapping may be so low that it does not significantly improve performance.

There are two aims to the current paper. The first is to introduce a performance criterion for Monte Carlo schemes of this general kind that differs in some interesting ways from traditional criteria, such as the magnitude of the subdominant eigenvalue of a related operator [11, 25]. More precisely, we use the theory of large deviations to define a "rate of convergence" for the empirical measure. The key observation here is that this rate, and hence the performance of parallel tempering, is monotonically increasing with respect to the swap rate a. Traditional wisdom in the application of parallel tempering has been that one should not attempt to swap too frequently. While an obvious reason is that the computational cost for swapping attempts might become a burden, it was also argued that frequent swapping would result in poor sampling. For a discussion on prior approaches to the question of how to set the swapping rate and an argument in favor of frequent swapping, see [20, 19].

The use of this large deviation criterion and the resulting monotonicity with respect to a directly suggest the second aim, which is to study parallel tempering in the limit as $a \to \infty$. Note that the computational cost due just to the swapping will increase without bound, even on bounded time intervals, when $a \to \infty$. However, we will construct an alternative scheme, which uses different process dynamics and a weighted empirical measure. Because this process no longer swaps particle positions, it and the weighted empirical measure have a well-defined limit as $a \to \infty$ which we call infinite swapping. In effect, the swapping is achieved through the proper choice of weights and state dependent diffusion coefficients. This is done for the case of both continuous and discrete time processes with multiple temperatures.

An outline of the paper is as follows. In the next section the swapping model in continuous time is introduced and the rate of convergence, as measured by a large deviations rate function, is defined. The alternative scheme which is equivalent to swapping but which has a well-defined limit is introduced, and its limit as $a \to \infty$ is identified. The following section considers the analogous limit model for more than two temperatures and discusses certain practical difficulties associated with direct implementation when the number of temperatures is not small. The continuous time model is used for illustration because both the large deviation rate and the weak limit of the appropriately redefined swapping model take a simpler form than those of discrete time models. However, the discrete time model is what is actually implemented in practice. To bridge the gap between continuous time diffusion models and discrete time models, in section 4 we discuss the idea of infinite swapping for continuous time Markov jump processes and prove that the key properties demonstrated for diffusion models hold here as well. We also state a uniform (in the swapping parameter) large deviation principle. The discrete time form actually used in numerical implementation is presented in section 5. Section 6 returns to the issue of implementation when the number of temperatures is not small. In particular, we resolve the difficulty of direct implementation of the infinite swapping models via approximation by what we call partial infinite swapping models. Section 7 gives numerical examples, and an appendix gives the proof of the uniform large deviation principle.

2. Diffusion models with two temperatures.

Although the implementation of parallel tempering uses a discrete time model, the motivation for the infinite swapping limit is best illustrated in the setting where the state process is a continuous time diffusion process. It is in this case that the large deviation rate function, as well as the construction of a process that is distributionally equivalent to the infinite swapping limit, is simplest. In order to minimize the notational overhead, we discuss in detail the two-temperature case. The extension to models with multiple temperatures is obvious and will be stated in the next section.

2.1. Model setup.

Let $(\bar{X}_1^a, \bar{X}_2^a)$ denote the Markov jump diffusion process of parallel tempering with swap rate a. That is, between swaps (or jumps), the process follows the diffusion dynamics (1.1). Jumps occur according to the state dependent intensity function $ag(\bar{X}_1^a, \bar{X}_2^a)$. At a jump time t, the particles swap locations, that is, $(\bar{X}_1^a(t), \bar{X}_2^a(t)) = (\bar{X}_2^a(t-), \bar{X}_1^a(t-))$. Hence for a smooth function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ the infinitesimal generator of the process is given by

$$\begin{aligned}\mathcal{L}^a f(x_1,x_2) = &- \langle \nabla_{x_1} f(x_1,x_2), \nabla V(x_1)\rangle - \langle \nabla_{x_2} f(x_1,x_2), \nabla V(x_2)\rangle \\ &+ \tau_1 \mathrm{tr}\left[\nabla^2_{x_1x_1} f(x_1,x_2)\right] + \tau_2 \mathrm{tr}\left[\nabla^2_{x_2x_2} f(x_1,x_2)\right] \\ &+ ag(x_1,x_2)\left[f(x_2,x_1) - f(x_1,x_2)\right],\end{aligned}$$

where $\nabla_{x_i} f$ and $\nabla^2_{x_ix_i} f$ denote the gradient and the Hessian matrix with respect to x_i, respectively, and tr denotes trace. Throughout the paper we also assume the growth condition

$$\lim_{r\to\infty} \inf_{x:|x|\geq r} \langle \nabla V(x), x/|x|\rangle = \infty. \tag{2.1}$$

This condition not only ensures the existence and uniqueness of the invariant distribution but also enforces the exponential tightness needed for the large deviation principle for the empirical measures.

Recall the definition of π in (1.2), and let μ be the corresponding Gibbs probability distribution, that is,

$$\mu(dx_1dx_2) = \pi(x_1,x_2)dx_1dx_2 \propto e^{-\frac{V(x_1)}{\tau_1}} e^{-\frac{V(x_2)}{\tau_2}} dx_1dx_2.$$

Straightforward calculations show that for any smooth function f which vanishes at infinity

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}^a f(x_1,x_2)\mu(dx_1dx_2) = 0. \tag{2.2}$$

Since the condition (2.1) implies that $V(x)\to\infty$ as $|x|\to\infty$, by Echeverria's theorem [5, Theorem 4.9.17], μ is the unique invariant probability distribution of the process $(\bar{X}^a_1, \bar{X}^a_2)$.

2.2. Rate of convergence by large deviations. It follows from the previous discussion and the ergodic theorem [1] that, for a fixed burn-in time B, with probability one (w.p.1)

$$\lambda_T^a \doteq \frac{1}{T}\int_B^{T+B} \delta_{(\bar{X}_1^a(t),\bar{X}_2^a(t))} dt \Rightarrow \mu$$

as $T \to \infty$. For notational simplicity we assume without loss of generality that $B = 0$ from now on. A basic question of interest is, How rapid is this convergence, and how does it depend on the swap rate a? In particular, what is the rate of convergence of the lower temperature marginal?

We note that standard measures one might use for the rate of convergence, such as the second eigenvalue of the associated operator, are not necessarily appropriate here. They provide only indirect information on the convergence properties of the empirical measure, which is the quantity of interest in the Monte Carlo approximation. Such measures properly characterize the convergence rate of the transition probability

$$p(x, dy; t) = P\left\{(\bar{X}_1^a(t), \bar{X}_2^a(t)) \in dy | (\bar{X}_1^a(0), \bar{X}_2^a(0)) = x\right\}, \quad x, y \in \mathbb{R}^d \times \mathbb{R}^d,$$

as $t \to \infty$. However, they neglect the time averaging effect of the empirical measure, an effect that is not present with the transition probability. In fact, it is easy to construct examples such as nearly periodic Markov chains for which the second eigenvalue suggests a slow convergence when in fact the empirical measure converges quickly [17].

Another commonly used criterion for algorithm performance is the notion of asymptotic variance [10, 12, 23]. For a given functional $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, one can establish a central limit theorem which asserts that as $T \to \infty$

$$\mathrm{Var}\left[\frac{1}{\sqrt{T}}\int_0^T f(\bar{X}_1^a(t), \bar{X}_2^a(t))\, dt\right] \to \sigma^2.$$

The magnitude of σ is used to measure the statistical efficiency of the algorithm. The asymptotic variance is closely related to the spectral properties of the underlying probability transition kernel [7, 17]. However, as with the second eigenvalue the usefulness of this criterion for evaluating performance of the empirical measure λ_T^a is not clear.

In this paper, we use the large deviation rate function to characterize the rate of convergence of a sequence of random probability measures. To be more precise, let S be a Polish space, that is, a complete and separable metric space. Denote by $\mathcal{P}(S)$ the space of all probability measures on S. We equip $\mathcal{P}(S)$ with the topology of weak convergence, though one can often use the stronger τ-topology [3]. Under the

weak topology, $\mathcal{P}(S)$ is metrizable and itself a Polish space. Note that the empirical measure λ_T^a is a random probability measure, that is, a random variable taking values in the space $\mathcal{P}(S)$.

DEFINITION 2.1. *A sequence of random probability measures $\{\gamma_T\}$ is said to satisfy a large deviation principle (LDP) with rate function $I : \mathcal{P}(S) \to [0, \infty]$ if for all open sets $O \subset \mathcal{P}(S)$,*

$$\liminf_{T\to\infty} \frac{1}{T} \log P\{\gamma_T \in O\} \geq - \inf_{\nu \in O} I(\nu),$$

for all closed sets $F \subset \mathcal{P}(S)$,

$$\limsup_{T\to\infty} \frac{1}{T} \log P\{\gamma_T \in F\} \leq - \inf_{\nu \in F} I(\nu),$$

and $\{\nu : I(\nu) \leq M\}$ is compact in $\mathcal{P}(S)$ for all $M < \infty$.

For our problem all rate functions encountered will vanish only at the unique invariant distribution μ and hence give information on the rate of convergence of λ_T^a. A larger rate function will indicate faster convergence, though this is an asymptotic statement valid only for sufficiently large T.

2.3. Explicit form of rate function. The large deviation theory for the empirical measure of a Markov process was first studied in [2]. Besides the Feller property, which will hold for all processes we consider, the validity of the LDP depends on two types of conditions. One is a so-called transitivity condition, which requires that there be times T_1 and T_2 such that for any $x, y \in \mathbb{R}^d \times \mathbb{R}^d$,

$$\int_0^{T_1} e^{-t} p(x, dz; t)dt \ll \int_0^{T_2} e^{-t} p(y, dz; t)dt,$$

where $\ll$ indicates that the measure in z on the left is absolutely continuous with respect to the measure on the right. For the jump diffusion process we consider here, this condition holds automatically since ∇V is bounded on bounded sets, g is bounded, and the diffusion coefficients are uniformly nondegenerate. The second type of condition is one that enforces a strong form of tightness, such as (2.1).

Under condition (2.1), the LDP holds for $\{\lambda_T^a : T > 0\}$ and the rate function, denoted by I^a, takes a fairly explicit form because the process is in continuous time and reversible [2, 13]. We will state the following result and omit the largely straightforward calculation since its role here is motivational. (A uniform LDP for the analogous jump Markov process will be stated in section 4, and its proof is given in the appendix.)

Let ν be a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ with smooth density. Define $\theta(x_1, x_2) \doteq [d\nu/d\mu](x_1, x_2)$. Then $I^a(\nu)$ can be expressed as

$$I^a(\nu) = J_0(\nu) + aJ_1(\nu), \tag{2.3}$$

where

$$J_0(\nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{8\theta(x_1,x_2)^2} \left[\tau_1 \left\|\nabla_{x_1}\theta(x_1,x_2)\right\|^2 + \tau_2 \left\|\nabla_{x_2}\theta(x_1,x_2)\right\|^2\right] \nu(dx_1 dx_2),$$

$$J_1(\nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x_1,x_2)\ell\left(\sqrt{\frac{\theta(x_2,x_1)}{\theta(x_1,x_2)}}\right) \nu(dx_1 dx_2),$$

and where $\ell(z) = z\log z - z + 1$ for $z \geq 0$ is familiar from the large deviation theory for jump processes.

The key observation is that the rate function $I^a(\nu)$ is affine in the swapping rate a, with $J_0(\nu)$ the rate function in the case of no swapping. Furthermore, $J_1(\nu) \geq 0$ with equality if and only if $\theta(x_2,x_1) = \theta(x_1,x_2)$ for ν-a.e. (x_1,x_2). This form of the rate function, and in particular its monotonicity in a, motivates the study of the *infinite swapping limit* as $a \to \infty$.

Remark 2.2. The limit of the rate function I^a satisfies

$$I^\infty(\nu) \doteq \lim_{a\to\infty} I^a(\nu) = \begin{cases} J_0(\nu) & \text{if } \theta(x_1,x_2) = \theta(x_2,x_1) \ \nu\text{-a.s.}, \\ \infty & \text{otherwise.} \end{cases}$$

Hence for $I^\infty(\nu)$ to be finite it is necessary that ν put exactly the same relative weight as μ on the points (x_1,x_2) and (x_2,x_1). Note that if a process could be constructed with I^∞ as its rate function, then with the large deviation rate as our criterion such a process improves on parallel tempering with finite swapping rate in exactly those situations where parallel tempering improves on the process with no swapping at all.

2.4. Infinite swapping limit. From a practical perspective, it may appear that there are limitations on how much benefit one obtains by letting $a \to \infty$. When implemented in discrete time, the overall jump intensity corresponds to the generation of roughly a independent random variables that are uniform on $[0,1]$ for each corresponding unit of continuous time, and based on each uniform variable a comparison is made to decide whether or not to make the swap. Hence even for fixed and finite T, the computations required to simulate a single trajectory scale like a as $a \to \infty$. Thus it is of interest if one can gain the benefit of the higher swapping rate without all the computational burden. This turns out to be possible but requires that we view the prelimit processes in a different way.

It is clear that the processes $(\bar{X}_1^a, \bar{X}_2^a)$ are not tight as $a \to \infty$ since the number of discontinuities of size $O(1)$ will grow without bound in any time interval of positive length. In order to obtain a limit, we consider alternative processes defined by

$$d\bar{Y}_1^a = -\nabla V(\bar{Y}_1^a)dt + \sqrt{2\tau_1 1_{\{\bar{Z}^a=0\}} + 2\tau_2 1_{\{\bar{Z}^a=1\}}}\, dW_1, \tag{2.4}$$

$$d\bar{Y}_2^a = -\nabla V(\bar{Y}_2^a)dt + \sqrt{2\tau_2 1_{\{\bar{Z}^a=0\}} + 2\tau_1 1_{\{\bar{Z}^a=1\}}}\, dW_2,$$

where $\bar{Z}^a$ is a jump process that switches from state 0 to state 1 with intensity $ag(\bar{Y}_1^a, \bar{Y}_2^a)$ and from state 1 to state 0 with intensity $ag(\bar{Y}_2^a, \bar{Y}_1^a)$. Compared to conventional parallel tempering, the processes $(\bar{Y}_1^a, \bar{Y}_2^a)$ swap the diffusion coefficients at the jump times rather than the physical locations of two particles with constant diffusion coefficients. For this reason, we refer to the solution to (2.4) as the *temperature swapped process*, in order to distinguish it from the *particle swapped process* $(\bar{X}_1^a, \bar{X}_2^a)$. We illustrate these processes in Figure 1. Note that the solid line and the dotted line represent $\bar{X}_1^a$ and $\bar{X}_2^a$, respectively. These processes have more and more frequent jumps of size $O(1)$ as $a \to \infty$. In contrast, the processes $(\bar{Y}_1^a, \bar{Y}_2^a)$ have varying diffusion coefficients. The figure attempts to also suggest features of the discrete time setting, with both successful and failed swap attempts.

Clearly the empirical measure of $(\bar{Y}_1^a, \bar{Y}_2^a)$ does not provide an approximation to μ. Instead, we should shift our attention between $(\bar{Y}_1^a, \bar{Y}_2^a)$ and $(\bar{Y}_2^a, \bar{Y}_1^a)$ depending on the value of $\bar{Z}^a$. Indeed, the random probability measures

$$\eta_T^a = \frac{1}{T}\int_0^T \left[1_{\{\bar{Z}^a(t)=0\}}\delta_{(\bar{Y}_1^a(t),\bar{Y}_2^a(t))} + 1_{\{\bar{Z}^a(t)=1\}}\delta_{(\bar{Y}_2^a(t),\bar{Y}_1^a(t))}\right] dt \tag{2.5}$$

have the same distribution as

$$\frac{1}{T}\int_0^T \delta_{(\bar{X}_1^a(s),\bar{X}_2^a(s))} ds$$

and hence converge to μ at the same rate. However, these processes and measures have well-defined limits in distribution as $a \to \infty$. More precisely, we have the following result. For the proof see [9]. Related (but more complex) calculations are needed to prove the uniform large deviation result given in Theorem 4.1.

THEOREM 2.3. *Assume that ∇V is locally Lipschitz continuous. Then for each T the sequence $(\bar{Y}_1^a, \bar{Y}_2^a, \eta_T^a)$ converges in distribution to $(\bar{Y}_1^\infty, \bar{Y}_2^\infty, \eta_T)$ as $a \to \infty$, where $(\bar{Y}_1^\infty, \bar{Y}_2^\infty)$ is the unique strong solution to*

$$d\bar{Y}_1^\infty = -\nabla V(\bar{Y}_1^\infty)dt + \sqrt{2\tau_1\rho(\bar{Y}_1^\infty, \bar{Y}_2^\infty) + 2\tau_2\rho(\bar{Y}_2^\infty, \bar{Y}_1^\infty)}dW_1, \tag{2.6}$$

$$d\bar{Y}_2^\infty = -\nabla V(\bar{Y}_2^\infty)dt + \sqrt{2\tau_2\rho(\bar{Y}_1^\infty, \bar{Y}_2^\infty) + 2\tau_1\rho(\bar{Y}_2^\infty, \bar{Y}_1^\infty)}dW_2,$$

(2.7)

$$\eta_T^\infty = \frac{1}{T}\int_0^T \left[\rho(\bar{Y}_1^\infty(t), \bar{Y}_2^\infty(t))\delta_{(\bar{Y}_1^\infty(t),\bar{Y}_2^\infty(t))} + \rho(\bar{Y}_2^\infty(t), \bar{Y}_1^\infty(t))\delta_{(\bar{Y}_2^\infty(t),\bar{Y}_1^\infty(t))}\right] dt,$$

and

$$\rho(x_1, x_2) \doteq \frac{\pi(x_1, x_2)}{\pi(x_2, x_1) + \pi(x_1, x_2)}.$$

The existence and form of the limit are due to the time scale separation between the fast $\bar{Z}^a$ process and the slow $(\bar{Y}_1^a, \bar{Y}_2^a)$ process. To give an intuitive explanation of the limit dynamics, consider the prelimit processes (2.4). Suppose that on a small time interval, the value of the slow process $(\bar{Y}_1^a, \bar{Y}_2^a)$ does not vary much, say $(\bar{Y}_1^a, \bar{Y}_2^a) \approx (x_1, x_2)$. Given the dynamics of the binary process $\bar{Z}^a$, it is easy to verify that as a tends to infinity the fractions of time that $\bar{Z}^a = 0$ and $\bar{Z}^a = 1$ are $\rho(x_1, x_2)$ and $\rho(x_2, x_1)$, respectively. This leads to the limit dynamics (2.6). When mapped back to the particle swapped process, $\rho(x_1, x_2)$ and $\rho(x_2, x_1)$ account for the fraction of time that $(\bar{X}_1^a, \bar{X}_2^a) = (x_1, x_2)$ and $(\bar{X}_1^a, \bar{X}_2^a) = (x_2, x_1)$, respectively, which naturally leads to the limit weighted empirical measure (2.7).

The weights ρ_1 and ρ_2 do not depend on the unknown normalization constant, and in fact

$$\rho(x_1, x_2) = \frac{e^{-\frac{V(x_1)}{\tau_1} - \frac{V(x_2)}{\tau_2}}}{e^{-\frac{V(x_1)}{\tau_1} - \frac{V(x_2)}{\tau_2}} + e^{-\frac{V(x_2)}{\tau_1} - \frac{V(x_1)}{\tau_2}}} \tag{2.8}$$

and

$$\rho(x_2, x_1) = 1 - \rho(x_1, x_2) = \frac{e^{-\frac{V(x_2)}{\tau_1} - \frac{V(x_1)}{\tau_2}}}{e^{-\frac{V(x_1)}{\tau_1} - \frac{V(x_2)}{\tau_2}} + e^{-\frac{V(x_2)}{\tau_1} - \frac{V(x_1)}{\tau_2}}}.$$

The following properties of the limit system are worth noting:

1. *Instantaneous equilibration of multiple locations.* Observe that the lower temperature component of this modified "empirical measure," i.e., the first marginal, uses contributions from both components at all times, corrected according to the weights. The form of the weights in (2.7) guarantees that the contributions to η_T^∞ from locations $(\bar{Y}_1^\infty, \bar{Y}_2^\infty)$ and $(\bar{Y}_2^\infty, \bar{Y}_1^\infty)$ are at any time perfectly balanced according to the invariant distribution on product space.

2. *Symmetry and invariant distribution.* While the marginals of η_T^∞ play very different roles, the dynamics of $\bar{Y}_1^\infty$ and $\bar{Y}_2^\infty$ are actually symmetric. Using Echeverria's theorem [5, Theorem 4.9.17], it can be shown that the unique invariant distribution of the process $(\bar{Y}_1^\infty, \bar{Y}_2^\infty)$ has the density

$$\frac{1}{2}[\pi(x_1, x_2) + \pi(x_2, x_1)].$$

   It then follows from the ergodic theorem that $\eta_T^\infty \Rightarrow \mu$ w.p.1 as $T \to \infty$. This is hardly surprising since μ is the invariant distribution for the prelimit processes $(\bar{X}_1^a, \bar{X}_2^a)$.

3. *Escape from local minima.* Finally it is worth commenting on the behavior of the diffusion coefficients as a function of the relative positions of $\bar{Y}_1^\infty$ and $\bar{Y}_2^\infty$ on an energy landscape. Recall that $\tau_1 < \tau_2$. Suppose that $\bar{Y}_1^\infty(t)$ is near the bottom of a local minimum (which for simplicity we set to be zero), while $\bar{Y}_2^\infty(t)$ is at a higher energy level, perhaps within the same local minimum. Then

$$\rho(y_1, y_2) \approx \frac{e^{-\frac{V(y_2)}{\tau_2}}}{e^{-\frac{V(y_2)}{\tau_2}} + e^{-\frac{V(y_2)}{\tau_1}}} \approx 1, \quad \rho(y_2, y_1) = 1 - \rho(y_1, y_2) \approx 0.$$

Thus to some degree the dynamics look like

$$d\bar{Y}_1^\infty = -\nabla V(\bar{Y}_1^\infty)dt + \sqrt{2\tau_1}dW_1,$$
$$d\bar{Y}_2^\infty = -\nabla V(\bar{Y}_2^\infty)dt + \sqrt{2\tau_2}dW_2;$$

i.e., the particle higher up on the energy landscape is given the greater diffusion coefficient, while the one near the bottom of the well is automatically given the lower coefficient. Hence the particle which is already closer to escaping from the well is automatically given the greater noise (within the confines of (τ_1, τ_2)). Recalling the role of the higher temperature particle is to more assiduously explore the landscape in parallel tempering; this is an interesting property.

One can apply results from [2] to show that the empirical measure of the infinite swapping limit $\{\eta_T^\infty : T > 0\}$ satisfies an LDP with rate function I^∞ as defined in Remark 2.2. However, to justify the claim that the infinite swapping model is truly superior to the finite swapping variant (note that $I^\infty \geq I^a$ for any finite a), one should establish a *uniform* LDP, which would show that I^∞ is the correct rate function for any sequence $\{a_T : T > 0\} \subset [0, \infty]$ such that $a_T \to \infty$ as $T \to \infty$. We omit the proof here, since in Theorem 4.1 the analogous result will be proved in the setting of continuous time jump Markov processes.

3. **Diffusion models with multiple temperatures.** In practice, parallel tempering uses swaps between more than two temperatures. A key reason is that if the gap between the temperatures is too large, then the probability of a successful swap under the (discrete time version of the) Metropolis rule (1.3) is far too low for the exchange of information to be effective. A natural generalization is to introduce, to the degree that computational feasibility is maintained, a ladder of higher temperatures and then attempt pairwise swaps between particles. There are a variety of schemes used to select which pair should attempt the swap, including deterministic and randomized rules for selecting only adjacent temperatures or an arbitrary pair of

temperatures. However, if one were to replace any of these particle swapped processes with its equivalent temperature swapped analogue and consider the infinite swapping limit, one would get the same system of process dynamics and weighted empirical measures which we now describe.

Suppose that besides the lowest temperature τ_1 (in many cases the temperature of principal interest), we introduce the collection of higher temperatures

$$\tau_1 < \tau_2 < \cdots < \tau_K.$$

Let $y = (y_1, y_2, \ldots, y_K) \in (\mathbb{R}^d)^K$ be a generic point in the state space of the process, and define a product Gibbs distribution with the density

$$\pi(y) = \pi(y_1, y_2, \ldots, y_K) \propto e^{-V(y_1)/\tau_1} e^{-V(y_2)/\tau_2} \cdots e^{-V(y_K)/\tau_K}.$$

The limit of the temperature swapped processes with K temperatures takes the form

$$d\bar{Y}_1^\infty = -\nabla V\left(\bar{Y}_1^\infty\right)dt + \sqrt{2\rho_{11}\tau_1 + 2\rho_{12}\tau_2 + \cdots + 2\rho_{1K}\tau_K}\,dW_1,$$
$$d\bar{Y}_2^\infty = -\nabla V\left(\bar{Y}_2^\infty\right)dt + \sqrt{2\rho_{21}\tau_1 + 2\rho_{22}\tau_2 + \cdots + 2\rho_{2K}\tau_K}\,dW_2,$$
$$\vdots$$
$$d\bar{Y}_K^\infty = -\nabla V\left(\bar{Y}_K^\infty\right)dt + \sqrt{2\rho_{K1}\tau_1 + 2\rho_{K2}\tau_2 + \cdots + 2\rho_{KK}\tau_K}\,dW_K.$$

To define these weights ρ_{ij} it is convenient to introduce some new notation.

Let S_K be the collection of all bijective mappings from $\{1,2,\ldots,K\}$ to itself. S_K has $K!$ elements, each of which corresponds to a unique permutation of the set $\{1,2,\ldots,K\}$, and S_K forms a group with the group action defined by composition. Let σ^{-1} denote the inverse of σ. Furthermore, for each $\sigma \in S_K$ and every $y = (y_1, y_2, \ldots, y_K) \in (\mathbb{R}^d)^K$, define $y_\sigma \doteq \left(y_{\sigma(1)}, y_{\sigma(2)}, \ldots, y_{\sigma(K)}\right)$.

At the level of the prelimit particle swapped process, we interpret the permutation σ to correspond to the event that the particles at location $y = (y_1, y_2, \ldots, y_K)$ are swapped to the new location $y_\sigma = \left(y_{\sigma(1)}, y_{\sigma(2)}, \ldots, y_{\sigma(K)}\right)$. Under the temperature swapped process, this corresponds to the event that particles initially assigned temperatures in the order $\tau_1, \tau_2, \ldots, \tau_K$ have now been assigned the temperatures $\tau_{\sigma^{-1}(1)}, \tau_{\sigma^{-1}(2)}, \ldots, \tau_{\sigma^{-1}(K)}$.

The identification of the infinite swapping limit of the temperature swapped processes is very similar to that of the two-temperature model in the previous section. By exploiting the time-scale separation, one can assume that in a small time interval the only motion is due to temperature swapping and the motion due to diffusion is negligible. Hence the fraction of time that the permutation σ is in effect should again be proportional to the relative weight assigned by the invariant distribution to y_σ, that is,

$$\pi(y_\sigma) = \pi\left(y_{\sigma(1)}, y_{\sigma(2)}, \ldots, y_{\sigma(K)}\right).$$

Thus if

$$w(y) \doteq \frac{\pi(y)}{\sum_{\theta \in S_K} \pi(y_\theta)},$$

then the fraction of time that the permutation σ is in effect is $w(y_\sigma)$. Note that for any y,

$$\sum_{\sigma \in S_K} w(y_\sigma) = 1.$$

Going back to the definition of the weights $\rho_{ij}(y)$, $i,j=1,\ldots,K$, it is clear that they represent the limit proportion of time that the ith particle is assigned the temperature j and hence will satisfy

$$\rho_{ij}(y) = \sum_{\sigma:\ \sigma(j)=i} w(y_\sigma).$$

Likewise the replacement for the empirical measure, accounting as it does for mapping the temperature swapped process back to the particle swapped process, is given by

$$\eta_T^\infty = \frac{1}{T}\int_0^T \sum_{\sigma\in S_K} w(\bar{Y}_\sigma^\infty(t))\delta_{\bar{Y}_\sigma^\infty(t)}dt, \tag{3.1}$$

where $\bar{Y}_\sigma^\infty(t) \doteq [\bar{Y}^\infty(t)]_\sigma = (\bar{Y}_{\sigma(1)}^\infty(t), \bar{Y}_{\sigma(2)}^\infty(t), \ldots, \bar{Y}_{\sigma(K)}^\infty(t))$.

The instantaneous equilibration property still holds for the infinite swapping system with multiple temperatures. That is, at any time $t\in[0,T]$ and given a current position $\bar{Y}^\infty(t)=y$, the weighted empirical measure η_T^∞ has contributions from all locations of the form y_σ, $\sigma\in S_K$, balanced exactly according to their relative contributions from the invariant density $\pi(y_\sigma)$. The dynamics of $\bar{Y}^\infty$ are again symmetric, and the density of the invariant distribution at point y is

$$\frac{1}{K!}\sum_{\sigma\in S_K}\pi(y_\sigma).$$

Remark 3.1. The infinite swapping process described above allows the most effective communication between all temperatures, and is the "best" in the sense that it leads to the largest large deviation rate function and hence the fastest rate of convergence. However, computation of the coefficients becomes very demanding for even moderate values of K since one needs to evaluate $K!$ terms from all possible permutations. In section 5 we discuss a more tractable and easily implementable family of schemes which are essentially approximations to the infinite swapping model presented in the current section and have very similar performance. We call the current model the full infinite swapping model since it uses the whole permutation group S_K, as opposed to the partial infinite swapping model in section 5 where only subgroups of S_K are used.

4. Infinite swapping for jump Markov processes. The continuous time diffusion model is a convenient vehicle for conveying the main idea of infinite swapping. In practice, however, algorithms are implemented in discrete time. In this section we discuss continuous time pure jump Markov processes and the associated infinite swapping limit. The purpose of this intermediate step is to serve as a bridge between the diffusion and discrete time Markov chain models. These two types of processes have some subtle differences regarding the infinite swapping limit, which is best illustrated through the continuous time jump Markov model.

In this section we discuss the two-temperature model and omit the completely analogous multiple-temperature counterpart. We will not refer to temperatures τ_1 and τ_2 to distinguish dynamics. Instead, let $\alpha_1(x, dy)$ and $\alpha_2(x, dy)$ be two probability transition kernels on $\mathbb{R}^d$ given $\mathbb{R}^d$. One can think of α_i as the dynamics under temperature τ_i for $i = 1, 2$. We assume that for each $i = 1, 2$ the stationary distribution μ_i associated with the transition kernel α_i admits the density π_i, in order to be consistent with the diffusion models, and define

$$\mu = \mu_1 \times \mu_2, \quad \pi(x_1, x_2) = \pi_1(x_1)\pi_2(x_2).$$

We assume that the kernels are Feller and have a density that is uniformly bounded with respect to Lebesgue measure. These conditions would always be satisfied in practice. Finally, we assume that the detailed balance or reversibility condition holds, that is,

$$\alpha_i(x, dz)\pi_i(x)dx = \alpha_i(z, dx)\pi_i(z)dz. \tag{4.1}$$

4.1. Model setup. In the absence of swapping (i.e., swapping rate $a = 0$), the dynamics of the system are as follows. Let $X^0 = \{X^0(t) = (X_1^0(t), X_2^0(t)) : t \geq 0\}$ denote a continuous time process taking values in $\mathbb{R}^d \times \mathbb{R}^d$. The probability transition kernel associated with the embedded Markov chain, denoted by $\bar{X}^0 = \{(\bar{X}_1^0(j), \bar{X}_2^0(j)) : j = 0, 1, \ldots\}$, is

$$P\{\bar{X}^0(j+1) \in (dy_1, dy_2) | \bar{X}^0(j) = (x_1, x_2)\} = \alpha_1(x_1, dy_1)\alpha_2(x_2, dy_2).$$

Without loss of generality, we assume that the jump times occur according to a Poisson process with rate one. In other words, let $\{\tau_i\}$ be a sequence of independent and identically distributed (i.i.d.) exponential random variables with rate one that are independent of $\bar{X}^0$. Then

$$X^0(t) = \bar{X}^0(j) \quad \text{for } \sum_{i=1}^{j} \tau_i \leq t < \sum_{i=1}^{j+1} \tau_i.$$

The infinitesimal generator of X^0 is such that for a given smooth function f,

$$\mathcal{L}^0 f(x_1, x_2) = \int_{\mathbb{R}^d \times \mathbb{R}^d} [f(y_1, y_2) - f(x_1, x_2)]\alpha_1(x_1, dy_1)\alpha_2(x_2, dy_2).$$

Owing to the detailed balance condition (4.1), the operator $\mathcal{L}^0$ is self-adjoint. Using arguments similar to but simpler than those used to prove the uniform LDP in Theorem 4.1, the large deviation rate function I^0 associated with the occupation measure

$$\eta_T^0 = \frac{1}{T} \int_0^T \delta_{X^0(t)}\, dt$$

can be explicitly identified: for any probability measure ν on $\mathbb{R}^d \times \mathbb{R}^d$ with $\nu \ll \mu$ and $\theta = d\nu/d\mu$,

$$I^0(\nu) = 1 - \int_{(\mathbb{R}^d \times \mathbb{R}^d)^2} \sqrt{\theta(x_1,x_2)\theta(y_1,y_2)}\pi(x_1,x_2)\alpha_1(x_1,dy_1)\alpha_2(x_2,dy_2)\,dx_1dx_2,$$

and I^0 is extended to all of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ by lower semicontinuous regularization.

4.2. Finite swapping model. Denote by $X^a = \{(X_1^a(t), X_2^a(t)) : t \geq 0\}$ the state process of the finite swapping model with swapping rate a, and let $\bar{X}^a = \{(\bar{X}_1^a(j), \bar{X}_2^a(j)) : j = 0,1,\ldots\}$ be the embedded Markov chain. The probability transition kernel for $\bar{X}^a$ is

$$P\{\bar{X}^a(j+1) \in (dy_1,dy_2) | \bar{X}^a(j) = (x_1,x_2)\} = \frac{1}{a+1}\alpha_1(x_1,dy_1)\alpha_2(x_2,dy_2)$$
$$+ \frac{a}{a+1}\left[g(x_1,x_2)\delta_{(x_2,x_1)}(dy_1,dy_2) + (1-g(x_1,x_2))\delta_{(x_1,x_2)}(dy_1,dy_2)\right],$$

where g is defined as in (1.3). Furthermore, let $\{\tau_i^a\}$ be a sequence of i.i.d. exponential random variables with rate $(a+1)$, and define

$$X^a(t) = \bar{X}^a(j) \quad \text{for } \sum_{i=1}^{j} \tau_i^a \leq t < \sum_{i=1}^{j+1} \tau_i^a.$$

In other words, the jumps occur according to a Poisson process with rate $a+1$. Note that there are two types of jumps. At any jump time, with probability $1/(a+1)$ it will be a jump according to the underlying probability transition kernels α_1 and α_2. With probability $a/(a+1)$ it will be an attempted swap which will succeed with the probability determined by g. As a grows, the swap attempts become more and more frequent. However, the time between two consecutive jumps of the first type will have the same distribution as

$$S^a = \sum_{i=1}^{N^a} \tau_i^a,$$

where N^a is a geometric random variable with parameter $1/(a+1)$. It is easy to argue that for any a the distribution of S^a is exponential with rate one. This observation will be useful when we derive the infinite swapping limit.

The infinitesimal generator $\mathcal{L}^a$ of $\bar{X}^a$ is such that for any smooth function f on $\mathbb{R}^d \times \mathbb{R}^d$,

$$\mathcal{L}^a f(x_1,x_2) = \int_{\mathbb{R}^d \times \mathbb{R}^d} [f(y_1,y_2) - f(x_1,x_2)]\alpha_1(x_1,dy_1)\alpha_2(x_2,dy_2)$$
$$+ ag(x_1,x_2)[f(x_2,x_1) - f(x_1,x_2)].$$

It is not difficult to check that the stationary distribution of $\bar{X}^a$ remains μ and that $\mathcal{L}^a$ is self-adjoint. As before, the large deviation rate function I^a for the occupation measure

(4.2) $$\eta_T^a = \frac{1}{T}\int_0^T \delta_{\boldsymbol{X}^a(t)}\, dt$$

can be explicitly identified. Indeed, for any probability measure ν on $\mathbb{R}^d \times \mathbb{R}^d$ with $\nu \ll \mu$ and $\theta = d\nu/d\mu$,

$$I^a(\nu) = I^0(\nu) + aJ(\nu),$$

where

$$J(\nu) = \int_{\mathbb{R}^d\times\mathbb{R}^d} g(x_1,x_2)\ell\left(\sqrt{\frac{\theta(x_2,x_1)}{\theta(x_1,x_2)}}\right)\nu(dx_1,dx_2).$$

Note that as before, I^a is monotonically increasing with respect to a. Since $J(\nu) \geq 0$ with equality if and only if $\theta(x_1,x_2) = \theta(x_2,x_1)$ ν-a.e., we have

(4.3) $$I^\infty(\nu) \doteq \lim_{a\to\infty} I^a(\nu) = \begin{cases} I^0(\nu) & \text{if } \theta(x_1,x_2) = \theta(x_2,x_1) \ \nu\text{-a.s.}, \\ \infty & \text{otherwise.} \end{cases}$$

4.3. Infinite swapping limit. The infinite swapping limit for $\boldsymbol{X}^a$ as $a \to \infty$ can be similarly obtained by considering the corresponding temperature swapped processes. Since the times between jumps determined by α_1 and α_2 are always exponential with rate one, the infinite swapping limit $\boldsymbol{Y}^\infty = (Y_1^\infty, Y_2^\infty)$ is a pure jump Markov process where jumps occur according to a Poisson process with rate one. In other words,

$$\boldsymbol{Y}^\infty(t) = \bar{\boldsymbol{Y}}^\infty(j) \quad \text{for } \sum_{i=1}^{j}\tau_i \leq t < \sum_{i=1}^{j+1}\tau_i,$$

where $\bar{\boldsymbol{Y}}^\infty$ is the embedded Markov chain and $\{\tau_i\}$ a sequence of i.i.d. exponential random variables with rate one. Furthermore, the probability transition kernel for $\bar{\boldsymbol{Y}}^\infty$ is

(4.4) $$\begin{aligned} &P\{\bar{\boldsymbol{Y}}^\infty(j+1) \in (dz_1,dz_2) | \bar{\boldsymbol{Y}}^\infty(j) = (y_1,y_2)\} \\ &\quad = \rho(y_1,y_2)\alpha_1(y_1,dz_1)\alpha_2(y_2,dz_2) + \rho(y_2,y_1)\alpha_2(y_1,dz_1)\alpha_1(y_2,dz_2), \end{aligned}$$

where the weight function ρ is defined as in Theorem 2.3. It is not difficult to argue that the stationary distribution for $\boldsymbol{Y}^\infty$ is

$$\bar{\mu}(dy_1,dy_2) = \frac{1}{2}[\pi(y_1,y_2) + \pi(y_2,y_1)]\, dy_1 dy_2,$$

and the weighted occupation measure

(4.5) $$\eta_T^\infty = \frac{1}{T}\int_0^T \left[\rho(Y_1^\infty(t),Y_2^\infty(t))\delta_{(Y_1^\infty(t),Y_2^\infty(t))} + \rho(Y_2^\infty(t),Y_1^\infty(t))\delta_{(Y_2^\infty(t),Y_1^\infty(t))}\right] dt$$

converges to $\mu(dx_1, dx_2) = \pi(x_1, x_2)dx_1dx_2$ as $T \to \infty$. It is obvious that the dynamics of the infinite swapping limit are symmetric and instantaneously equilibrate the contribution from (Y_1, Y_2) and (Y_2, Y_1) according to the invariant measure, owing to the weight function ρ.

We have the following uniform LDP result, which justifies the superiority of the infinite swapping model. Its proof is deferred to the appendix. It should be noted that rate identification is not covered by the existing literature, even in the case of a fixed swapping rate, due to the pure jump nature of the process.

THEOREM 4.1. *The occupation measure* $\{\eta_T^\infty : T > 0\}$ *satisfies an LDP with rate function* I^∞. *More generally, define the finite swapping model as in subsection* 4.2. *Consider any sequence* $\{a_T : T > 0\} \subset [0, \infty]$ *such that* $a_T \to \infty$ *as* $T \to \infty$, *and interpret* $a_T < \infty$ *to mean that* $\eta_T^{a_T}$ *is defined by* (4.2) *with* $a = a_T$, *and* $a_T = \infty$ *to mean that* $\eta_T^{a_T}$ *is defined by* (4.5). *Then* $\{\eta_T^{a_T} : T > 0\}$ *satisfies an LDP with the rate function* I^∞ *defined in* (4.3).

5. Discrete time process models.

5.1. Conventional parallel tempering algorithms. In the discrete time, multitemperature algorithms that are actually implemented, a swap is attempted after a deterministic or random number of time steps, with a success probability of the form (1.3). The two temperatures corresponding to particles for which a swap is attempted can be chosen according to a deterministic or random schedule, and as noted previously they are usually adjacent since otherwise the success probability (1.3) will be too small to allow efficient exchange of information.

As before it suffices to describe the algorithm in the setting of two temperatures. As in section 4, let $\alpha_i(x, dy)$ denote the probability transition kernel for temperature τ_i whose stationary distribution has a density π_i for $i = 1, 2$. For now let $N = 1/a$ be a fixed positive integer that determines the frequency of swap attempts. Let $\bar{X} = \{(\bar{X}_1(j), \bar{X}_2(j)) : j = 0, 1, \ldots\}$ denote the state process. Then the evolution of the dynamics is as follows. For any integer $k \geq 1$ and $(k-1)(N+1) \leq j \leq k(N+1)-2$,

$$P\{\bar{X}(j+1) \in (dy_1, dy_2) | \bar{X}(j) = (x_1, x_2)\} = \alpha_1(x_1, dy_1)\alpha_2(x_2, dy_2),$$

and for $j = k(N+1) - 1$,

$$P\{\bar{X}(j+1) = (x_2, x_1) | \bar{X}(j) = (x_1, x_2)\} = g(x_1, x_2),$$
$$P\{\bar{X}(j+1) = (x_1, x_2) | \bar{X}(j) = (x_1, x_2)\} = 1 - g(x_1, x_2).$$

Thus a swap is attempted after every N ordinary time steps based on the underlying transition kernels α_1 and α_2. The case $N = 1/a$ with a an integer greater than one corresponds to the case where multiple swaps are attempted between two ordinary time steps. The unique invariant distribution of $\bar{X}$ is $\mu(dx_1dx_2) = \pi(x_1, x_2)\, dx_1dx_2$, regardless of the value of N, and the occupation measure

$$\frac{1}{J}\sum_{j=0}^{J-1} \delta_{\boldsymbol{X}(j)}$$

converges to μ as $J \to \infty$ a.s.

Remark 5.1. Note that N could be random. For example, if N is chosen to be a geometric random variable with mean λ, then $\bar{X}$ is exactly the embedded Markov chain of the pure jump Markov process $\bar{X}^a$ with $a = 1/\lambda$ in subsection 4.2.

5.2. Infinite swapping model. As with the continuous time case, to produce a well-defined limit one must consider the temperature swapped process and then consider the limit as swapping frequency tends to infinity. It turns out that the limit is exactly the embedded Markov chain for the pure jump Markov process in subsection 4.3. That is, the infinite swapping limit in discrete time is a Markov chain $\bar{Y}^\infty = \{(\bar{Y}_1^\infty(j), \bar{Y}_2^\infty(j)) : j = 0, 1, \ldots\}$ with the transition kernel

$$\rho(y_1, y_2)\alpha_1(y_1, dz_1)\alpha_2(y_2, dz_2) + \rho(y_2, y_1)\alpha_2(y_1, dz_1)\alpha_1(y_2, dz_2). \tag{5.1}$$

The corresponding weighted empirical measure is

(5.2)

$$\eta_J^\infty \doteq \frac{1}{J}\sum_{j=0}^{J-1}\left[\rho(\bar{Y}_1^\infty(j), \bar{Y}_2^\infty(j))\delta_{(\bar{Y}_1^\infty(j), \bar{Y}_2^\infty(j))} + \rho(\bar{Y}_2^\infty(j), \bar{Y}_1^\infty(j))\delta_{(\bar{Y}_2^\infty(j), \bar{Y}_1^\infty(j))}\right].$$

The generalization to multiple temperatures is also straightforward. Suppose that there are K temperatures. Denote the infinite swapping limit process by $\bar{Y} = \{\bar{Y}(j) : j = 0, 1, \ldots\}$, which is a Markov chain taking values in the space $(\mathbb{R}^d)^K$. Given that the current state of the chain is $\bar{Y}^\infty(j) = y = (y_1, \ldots, y_K)$, define as before the weights

$$w(y) \doteq \frac{\pi(y)}{\sum_{\theta \in S_K} \pi(y_\theta)}.$$

Then the transition kernel of $\bar{Y}^\infty$ is

$$P(\bar{Y}^\infty(j+1) \in dz | \bar{Y}^\infty(j) = y) = \sum_{\sigma \in S_K} w(y_\sigma)\alpha_1(y_{\sigma(1)}, dz_{\sigma(1)}) \cdots \alpha_K(y_{\sigma(K)}, dz_{\sigma(K)}).$$

The discrete time numerical approximation to the invariant distribution is

$$\eta_J^\infty \doteq \frac{1}{J}\sum_{j=0}^{J-1}\sum_{\sigma \in S_K} w(\bar{Y}_\sigma^\infty(j))\delta_{\bar{Y}_\sigma^\infty(j)}.$$

Remark 5.2. It is not difficult to derive LDPs for the discrete time finite swapping or infinite swapping models. However, it remains an open question whether the rate function is monotonic with respect to the swap rate (frequency). However, the discrete time large deviation rate function can be obtained from that of the continuous time pure jump Markov process models through the contraction principle, and the two coincide in the limit as the transition kernels α_i correspond to an infinitesimal time step for the diffusion process (1.1). Hence the discrete time rate function will be at least approximately monotone, and in this sense the infinite swapping limit should (at least approximately) dominate all finite swapping algorithms. This is supported by the data presented in section 7 and the much more extensive empirical study presented in [14].

6. Partial infinite swapping. As noted in section 3, the number of weights and their calculation can become unwieldy for infinite swapping even when the number of temperatures is moderate. In this section we construct algorithms that maintain most of the benefit of the infinite swapping algorithm but at a much lower computational cost. In the first subsection we describe the infinite swapping limit models when only a subgroup of the permutations of the particles (respectively, temperatures) are allowed by the prelimit particle (respectively, temperature) swapped process. The computational complexity of these limit models will be controlled by limiting the number of permutations that communicate with each other through the swapping mechanism. The infinite swapping models in this subsection will be called *partial infinite swapping models*, as opposed to the full infinite swapping models in the previous section. The second subsection shows how such partial infinite swapping schemes can be interwoven to approximate the full infinite swapping model.

6.1. Partial infinite swapping models. We consider subsets A of S_K with the property that A is an algebraic subgroup of S_K. That is, the following hold:

1. the identity belongs to A;
2. if $\sigma_1, \sigma_2 \in A$, then $\sigma_1 \circ \sigma_2 \in A$, where $\circ$ denotes composition;
3. if $\sigma \in A$, then $\sigma^{-1} \in A$.

Although one can write down a partial infinite swapping model that corresponds to instantaneous equilibration for an arbitrary subset A, it is only when A is a subgroup that the corresponding partial infinite swapping process has an interpretation as the limit of parallel tempering–type processes. When alternating between partial infinite swapping processes, a "handoff" rule will be needed, and it is only for those which correspond to subgroups that such a handoff rule is well defined. This point is discussed in some detail in the next section.

The definition of the partial infinite swapping process based on A is completely analogous to that of the full infinite swapping process. The state process $\{\bar{Y}(j) : j = 0, 1, \ldots\}$ is a Markov chain with the transition kernel

$$\alpha^A(y, dz) \doteq \sum_{\sigma \in A} \tilde{w}^A(y_\sigma)\, \alpha_1(y_{\sigma(1)}, dz_{\sigma(1)}) \cdots \alpha_K(y_{\sigma(K)}, dz_{\sigma(K)}), \tag{6.1}$$

and the weighted empirical measure is

$$\tilde{\eta}_J \doteq \frac{1}{J}\sum_{j=0}^{J-1}\sum_{\sigma\in A}\tilde{w}^A(\bar{Y}_\sigma(j))\delta_{\bar{Y}_\sigma(j)},$$

where the weight function $\tilde{w}^A$ is defined by

$$\tilde{w}^A(y) \doteq \frac{\pi(y)}{\sum_{\theta\in A}\pi(y_\theta)}, \tag{6.2}$$

and satisfies for any y

$$\sum_{\sigma\in A}\tilde{w}^A(y_\sigma) = 1.$$

We omit the dependence on both $a = \infty$ and A from the notation. Note that in contrast with the full swapping system, it is only those permutations of y corresponding to $\sigma \in A$ that are balanced according to the invariant distribution in their contributions to $\tilde{\eta}_J$.

To illustrate the construction we present a few examples. With a standard abuse of notation denote the permutation σ such that $\sigma(i) = a_i$ by the form $(a_1, a_2, \ldots, a_K)$. In particular, $(1, 2, \ldots, K)$ is the identity of the group S_K.

EXAMPLE 6.1. *Let $K = 4$ and $A = \{(1,2,3,4),(2,1,3,4)\}$. This corresponds to only allowing swaps between temperatures τ_1 and τ_2 at the prelimit. Define*

$$\tilde{w}(y) = \frac{\pi(y_1,y_2,y_3,y_4)}{\pi(y_1,y_2,y_3,y_4)+\pi(y_2,y_1,y_3,y_4)}.$$

The probability transition kernel of the corresponding partial infinite swapping process is given by

$$\begin{aligned}&\tilde{w}(y_1,y_2,y_3,y_4)\,\alpha_1(y_1,dz_1)\alpha_2(y_2,dz_2)\alpha_3(y_3,dz_3)\alpha_4(y_4,dz_4)\\&\quad+\tilde{w}(y_2,y_1,y_3,y_4)\,\alpha_1(y_2,dz_2)\alpha_2(y_1,dz_1)\alpha_3(y_3,dz_3)\alpha_4(y_4,dz_4),\end{aligned}$$

and the contribution to the weighted empirical measure is

$$\tilde{w}(\bar{Y}_1,\bar{Y}_2,\bar{Y}_3,\bar{Y}_4)\,\delta_{(\bar{Y}_1,\bar{Y}_2,\bar{Y}_3,\bar{Y}_4)} + \tilde{w}(\bar{Y}_2,\bar{Y}_1,\bar{Y}_3,\bar{Y}_4)\,\delta_{(\bar{Y}_2,\bar{Y}_1,\bar{Y}_3,\bar{Y}_4)}.$$

Note that with π_{ij} denoting the marginal invariant distribution on the ith and jth components, the weight function can be written as

$$\tilde{w}(y) = \frac{\pi_{12}(y_1,y_2)}{\pi_{12}(y_1,y_2)+\pi_{12}(y_2,y_1)},$$

which is consistent with the weights of the two-temperature model in Theorem 2.3.

EXAMPLE 6.2. *Again take $K = 4$, but this time use the subgroup generated by $(2,1,3,4)$ and $(1,2,4,3)$, i.e., $A = \{(1,2,3,4),(2,1,3,4),(1,2,4,3),(2,1,4,3)\}$. Then the dynamics are given by*

$$\begin{aligned}
&\tilde{w}(y_1,y_2,y_3,y_4)\,\alpha_1(y_1,dz_1)\alpha_2(y_2,dz_2)\alpha_3(y_3,dz_3)\alpha_4(y_4,dz_4)\\
&\quad+\tilde{w}(y_2,y_1,y_3,y_4)\,\alpha_1(y_2,dz_2)\alpha_2(y_1,dz_1)\alpha_3(y_3,dz_3)\alpha_4(y_4,dz_4)\\
&\qquad+\tilde{w}(y_1,y_2,y_4,y_3)\,\alpha_1(y_1,dz_1)\alpha_2(y_2,dz_2)\alpha_3(y_4,dz_4)\alpha_4(y_3,dz_3)\\
&\qquad\quad+\tilde{w}(y_2,y_1,y_4,y_3)\,\alpha_1(y_2,dz_2)\alpha_2(y_1,dz_1)\alpha_3(y_4,dz_4)\alpha_4(y_3,dz_3),
\end{aligned}$$

where the weight function $\tilde{w}$ is defined by

$$\tilde{w}(y) = \frac{\pi(y_1,y_2,y_3,y_4)}{\pi(y_1,y_2,y_3,y_4)+\pi(y_2,y_1,y_3,y_4)+\pi(y_1,y_2,y_4,y_3)+\pi(y_2,y_1,y_4,y_3)}.$$

The contribution to the weighted empirical measure is

$$\begin{aligned}
&\tilde{w}(\bar{Y}_1,\bar{Y}_2,\bar{Y}_3,\bar{Y}_4)\,\delta_{(\bar{Y}_1,\bar{Y}_2,\bar{Y}_3,\bar{Y}_4)}+\tilde{w}(\bar{Y}_2,\bar{Y}_1,\bar{Y}_3,\bar{Y}_4)\,\delta_{(\bar{Y}_2,\bar{Y}_1,\bar{Y}_3,\bar{Y}_4)}\\
&\quad+\tilde{w}(\bar{Y}_1,\bar{Y}_2,\bar{Y}_4,\bar{Y}_3)\,\delta_{(\bar{Y}_1,\bar{Y}_2,\bar{Y}_4,\bar{Y}_3)}+\tilde{w}(\bar{Y}_2,\bar{Y}_1,\bar{Y}_4,\bar{Y}_3)\,\delta_{(\bar{Y}_2,\bar{Y}_1,\bar{Y}_4,\bar{Y}_3)}.
\end{aligned}$$

EXAMPLE 6.3. *We let $K = 3$ and take A to be the subgroup of S_K generated by the rotation $(2,3,1)$, i.e., $A = \{(1,2,3),(2,3,1),(3,1,2)\}$. Then the dynamics are given by*

$$\begin{aligned}
&\tilde{w}(y_1,y_2,y_3)\,\alpha_1(y_1,dz_1)\alpha_2(y_2,dz_2)\alpha_3(y_3,dz_3)\\
&\quad+\tilde{w}(y_2,y_3,y_1)\,\alpha_1(y_2,dz_2)\alpha_2(y_3,dz_3)\alpha_3(y_1,dz_1)\\
&\qquad+\tilde{w}(y_3,y_1,y_2)\,\alpha_1(y_3,dz_3)\alpha_2(y_1,dz_1)\alpha_3(y_2,dz_2),
\end{aligned}$$

where

$$\tilde{w}(y) = \frac{\pi(y_1,y_2,y_3)}{\pi(y_1,y_2,y_3)+\pi(y_2,y_3,y_1)+\pi(y_3,y_1,y_2)}$$

and the contribution to the weighted empirical measure is

$$\tilde{w}_1(\bar{Y}_1,\bar{Y}_2,\bar{Y}_3)\,\delta_{(\bar{Y}_1,\bar{Y}_2,\bar{Y}_3)}+\tilde{w}_2(\bar{Y}_2,\bar{Y}_3,\bar{Y}_1)\delta_{(\bar{Y}_2,\bar{Y}_3,\bar{Y}_1)}+\tilde{w}_3(\bar{Y}_3,\bar{Y}_1,\bar{Y}_3)\,\delta_{(\bar{Y}_3,\bar{Y}_1,\bar{Y}_3)}.$$

The first two examples would correspond to the infinite swapping limit of a standard parallel tempering process, where swaps between only 1 and 2 are allowed in the first example, and swaps between 1 and 2 and swaps between 3 and 4 are allowed in the second. Note that the computational complexity does not increase significantly between the first and second examples. The third example corresponds

to a very different sort of prelimit process, in which "rotations" of the coordinates $(y_1, y_2, y_3) \to (y_2, y_3, y_1) \to (y_3, y_1, y_2) \to (y_1, y_2, y_3)$ are allowed. One can devise a Metropolis-type rule that allows such "swaps" and yields the indicated infinite swapping system.

6.2. Approximating full infinite swapping by partial swapping. In this section we consider the issue of alternating between such partial infinite swapping systems to approximate the full infinite swapping limit. Let A and B be subgroups of S_K. A and B are said to *generate* S_K if the smallest subgroup that contains A and B itself is S_K. Note that the total number of permutations in $A \cup B$ can be significantly

smaller than $K!$, the size of S_K. In fact, it is possible to construct subgroups A and B that generate S_K and that the total number of permutations in $A \cup B$ is of order K. There is an obvious extension to more than two subgroups.

EXAMPLE 6.4. *Let $K = 4$, and let A be generated by $\{(2,1,3,4),(1,3,2,4)\}$ and B be generated by $\{(1,3,2,4),(1,2,4,3)\}$, respectively. Thus A is the collection of six permutations that fix the last component and allow all rearrangements of the first three, while B fixes the first component and allows all rearrangements of the last three. Then A and B generate S_K.*

EXAMPLE 6.5. *Let $K = 4$, and let A and B be subgroups generated by $\{(2,1,3,4)\}$ and $\{(2,3,4,1)\}$, respectively. In other words, $A = \{(1,2,3,4),(2,1,3,4)\}$ corresponds to allowing permutations only between the first two components, while $B = \{(1,2,3,4),(2,3,4,1),(3,4,1,2),(4,1,2,3)\}$ corresponds to cycling of the four temperatures. Then A and B generate S_K.*

To keep the computational cost controlled, one can approximate the full infinite swapping model by alternating between partial infinite swapping processes whose associated subgroups generate the whole group. However, one must be careful in how the "handoff" is made when switching between different partial swapping models. It turns out that one cannot simply switch between different partial infinite swapping dynamics (i.e., transition kernels). Recall that in order to get a consistent approximation to the desired target invariant distribution we do not use the empirical measure generated by $\bar{Y}$ but rather a carefully constructed weighted empirical measure that works with several permutations of $\bar{Y}$. Simply switching the dynamics and weights will in fact produce an algorithm that may not converge to the target distribution.

To see how one should design a handoff rule, note that if one considers a collection of transition kernels each having the same invariant distribution and alternates between them in a way that does not depend on the outcomes prior to a switch, then the resulting empirical measure will in fact converge to the common invariant distribution. This fact is used (at least implicitly) in the parallel tempering algorithm itself, where one alternates the pair of particles being considered for swapping according to deterministic or random rules so long as the random rules do not rely on previously observed outcomes.

Now we use the fact that each partial infinite swapping model is a limit of either a parallel tempering algorithm where only some pairs of particles are considered for swapping or some more general form of parallel tempering which would allow groups of particles to simultaneously swap (according to an appropriate Metropolis-type acceptance rule). An example in the earlier category would be A in Example 6.4, which arises if only the pairs corresponding to temperatures τ_1, τ_2 and τ_2, τ_3 are allowed to swap, whereas an example in the latter category would be B in Example 6.5, which corresponds to allowing the particle at temperature τ_i to move to the location of the particle at temperature τ_{i-1} (with $\tau_0 = \tau_4$), and vice versa. Furthermore, each of these partial infinite swapping models arises as a limit of transition kernels of the corresponding temperature swapped processes which preserve the same common invariant distribution. In taking the limit as the swap rate tends to infinity, the correspondence between particle locations for a particle swapped process and the "instantaneously equilibrated" temperature swapped process $\bar{Y}$ is lost. However, one can construct a consistent algorithm by *reconstructing this correspondence.* In fact one should choose the particle location according to the probabilities (under the invariant distribution) associated with the various permutations in the subgroup. See subsection 6.3 for a more detailed discussion on the intuition behind this approximation algorithm.

We next present an algorithm for alternating between two partial infinite swapping dynamics. The restriction to two is for notational convenience only. Suppose that the dynamics are indexed by the corresponding subgroups A and B and that n_A steps of subgroup A are to be alternated with n_B steps of subgroup B. For simplicity we do not describe a "burn-in" period. As in (6.1) and (6.2) we let $\alpha^A(y, dz)$ and $\tilde{w}^A(y)$ denote the transition kernel for A and the weights allocated to the permutation $\sigma \in A$, respectively, and similarly for B.

ALGORITHM 6.6 (approximation to full infinite swapping).

1. *Initialization:* $\bar{X}^A(0) = \bar{Y}(0;0) \in (\mathbb{R}^d)^K$, $\ell = 1$.
2. *Loop* ℓ*:*
   (a) *Initialization for A dynamics: set* $\bar{Y}(\ell;0) = \bar{X}^A(\ell-1)$.
   (b) *Subgroup A dynamics: update* $\bar{Y}(\ell;k)$, $k = 1, \ldots, n_A$, *according to the transition kernel* α^A, *and add*

$$\sum_{\sigma \in A} \tilde{w}^A(\bar{Y}_\sigma(\ell;k))\delta_{\bar{Y}_\sigma(\ell;k)}$$

   to the unnormalized empirical measure.
   (c) *Reconstructing particle locations at the end of A dynamics: Let* $\bar{X}^B(\ell)$ *be a random sample from the set* $\{\bar{Y}_\sigma(\ell;n_A) : \sigma \in A\}$ *according to the weights* $\{\tilde{w}^A(\bar{Y}_\sigma(\ell;n_A)) : \sigma \in A\}$.
   (d) *Initialization for B dynamics: set* $\bar{Y}(\ell;n_A) = \bar{X}^B(\ell)$.
   (e) *Subgroup B dynamics: update* $\bar{Y}(\ell;k)$, $k = n_A+1, \ldots, n_A+n_B$, *according to the transition kernel* α^B, *and add*

$$\sum_{\sigma \in B} \tilde{w}^B(\bar{Y}_\sigma(\ell;k))\delta_{\bar{Y}_\sigma(\ell;k)}$$

to the unnormalized empirical measure.

(f) *Reconstructing particle locations at the end of B dynamics: Let* $\bar{X}^A(\ell)$ *be a random sample from the set* $\{\bar{Y}_\sigma(\ell;n_A+n_B):\sigma\in B\}$ *according to the weights* $\{\tilde{w}^B(\bar{Y}_\sigma(\ell;n_A+n_B)):\sigma\in B\}$.

(g) *Set* $\ell=\ell+1$, *and loop back to* (a).

3. *Normalize the empirical measure.*

6.3. Discussions on the approximation. In this section we further discuss the intuition underlying the handoff rule between different partial infinite swapping dynamics and the approximation algorithm of the previous section. We temporarily assume that the model is in continuous time since the intuition is most transparent in this case.

For simplicity let us assume that there are three temperatures and two groups $A=\{(1,2,3),(2,1,3)\}$ and $B=\{(1,2,3),(1,3,2)\}$. That is, under group A dynamics, only pairwise swaps between the temperatures τ_1 and τ_2 are allowed, while under the group B dynamics, only the swaps between τ_2 and τ_3 are allowed. See Figure 2.

Consider the following prelimit swapping model. Let the swap rate be a. The dynamics corresponding to group A and group B will be alternated on time intervals of length h. Hence on the interval $[2kh,(2k+1)h)$ the particle swapped process $(\bar{X}_1^a,\bar{X}_2^a,\bar{X}_3^a)$ involves swaps only between temperatures τ_1 and τ_2. One can easily construct the corresponding temperature swapped process $(\bar{Y}_1^a,\bar{Y}_2^a,\bar{Y}_3^a)$ as before. Note that $\bar{X}_3^a=\bar{Y}_3^a$ on this time interval. Similarly, on the interval $[(2k+1)h,(2k+2)h)$, only swaps between τ_2 and τ_3 are allowed and on this interval $\bar{X}_1^a=\bar{Y}_1^a$. Note that there is no ambiguity for the prelimit processes at the switch times $t=h,2h,\ldots$ since the locations of the particles $(\bar{X}_1^a,\bar{X}_2^a,\bar{X}_3^a)$ are known.

Now consider the limit as $a\to\infty$ with h being fixed. Without loss of generality, we will discuss only how to deal with the switch of the dynamics at time $t=h$. On the time interval $[0,h)$ we have the partial infinite swapping limit process $\bar{Y}^A=(\bar{Y}_1,\bar{Y}_2,\bar{Y}_3)$ that corresponds to the group A. Similarly it is clear that on the time interval $[h,2h)$ we should have the partial infinite swapping process $\bar{Y}^B$ corresponding to the group B. The problem is, however, by taking the limit, we lose the information on the locations of the particles $(\bar{X}_1^a,\bar{X}_2^a,\bar{X}_3^a)$. Unless we can somehow recover this information at the switch time $t=h$ to assign $\bar{Y}^B(h)$, we cannot determine the dynamics of $\bar{Y}^B$ on $[h,2h)$. The key is to recall that the infinite swapping limit instantaneously equilibrates multiple locations according to the invariant distribution. In other words, given $\bar{Y}^A(h-)=y=(y_1,y_2,y_3)$, the locations of the particles are distributed according to

$$\sum_{\sigma\in A}\tilde{w}^A(y_\sigma)\delta_{y_\sigma}=\frac{\pi(y_1,y_2,y_3)\delta_{(y_1,y_2,y_3)}+\pi(y_2,y_1,y_3)\delta_{(y_2,y_1,y_3)}}{\pi(y_1,y_2,y_3)+\pi(y_2,y_1,y_3)}.$$

Therefore, in order to identify the locations of the particles at time h, we will take a random sample from this distribution once $\bar{Y}^A(h-)$ is known. This explains the handoff rule used at the switch times of partial infinite swapping processes.

Now we let $h \to 0$. Since A and B generate the whole permutation group S_K, it is easy to check that at each time instant, the locations of $\{y_\sigma : \sigma \in S_K\}$ are equilibrated according to their invariant distribution, and therefore in the limit we will attain the full infinite swapping model. This can be made rigorous by exploiting the time scale separation between the slow diffusion processes $(\bar{Y}^A, \bar{Y}^B)$ and the fast switching process. We omit the proof because the discussion is largely motivational.

Coming back to the discrete time partial infinite swapping model, it is clear that Algorithm 6.6 is nothing but a straightforward adaption of the preceding discussion to discrete time. The only difference is that one cannot establish an analogous result regarding approximation to the full infinite swapping model as in continuous time. The subtlety here is that in continuous time, as $h \to 0$, one can basically ignore any effect from the diffusion on any small time interval and assume that the process is making jumps only between different permutations of a fixed triple (y_1, y_2, y_3). This time scale separation is no longer valid in discrete time. In this setting, the performance of a scheme based on interweaving partial infinite swapping schemes lies between parallel tempering and full infinite swapping, and computational results suggest that it is closer to the latter than the former.

The issue of which interwoven partial schemes will perform best is an open question. In practice we have used schemes of the following form. Suppose that a set of say 45 temperatures is given. We then partition 45 into blocks of sizes $3, 6, \ldots, 6$, with the first block containing the lowest three temperatures, the second block the next six, and so on. Dynamic A then is given by allowing all permutations within each block. Note that the complexity of the coefficients is then no worse than 6!. In dynamic B we use the partition $6, 6, \ldots, 6, 3$. The form of the partial scheme is heuristically motivated by allowing the largest possible overlap between the different blocks when switching between dynamics, subject to the constraint that blocks be of size no greater than 6.

7. Numerical examples. In this section we present data comparing parallel tempering at various swap rates and both full and partial infinite swapping. We present what we call "relaxation studies." The quantity of interest is the average potential energy of the lowest temperature component under the invariant distribution. In these studies, the system is run a long time to reach equilibrium, after which it is repeatedly pushed out of equilibrium and we measure the time needed to "relax" back to equilibrium. Each cycle consists of temporarily raising the temperatures of some of the lowest temperature components for a number of steps sufficient to push the average potential energy away from the "true" value (as measured by either sample or time averages). The temperatures are then returned to their "true" values for a fixed number of steps, and the process is then repeated 2,000 times. We plot the average of the 2,000 samples as a function of the number of moves, and the performance of the algorithm is captured by the rate at which these averages approach the correct value.

Figures 3 and 4 present data for a Lennard-Jones cluster of 13 atoms, using the "smart Monte Carlo" scheme of [18] for the simulation of the dynamics, which produces a relatively large move in configuration space for each step. The "true" value is approximately -42.92. This is a relatively simple model and was studied using only four temperatures. The temperatures are dropped to the true values at step 50. Infinite swapping converges more rapidly than any of the parallel tempering schemes. We see in Figure 3 that the most efficient of the parallel tempering schemes appears to use an attempted swap rate of around 64%. (The rates that would typically be used in such calculations are in the range of 5–10%.) Figure 4 magnifies a portion of the graph, but it plots only the best parallel tempering result and adds a partial infinite swapping result based on blocks of the form 1,3 and 3,1, and with a handoff at each Metropolis step. Little difference is observed between the partial and full forms, though exclusive use of either of the partial forms by itself performs poorly.

The Lennard-Jones cluster of 13 atoms is not a particularly demanding problem, but it is presented so a comparison can be made between the full and partial infinite swapping forms. A much more complex example is the Lennard-Jones cluster of 38 atoms. Data for this example obtained using a 45-temperature ensemble is given in Figure 5. Because full infinite swapping is impossible for this larger computational ensemble, we use the partial form. For comparison, results are also presented for parallel tempering.

The details concerning the computational methods underlying both the parallel tempering and infinite swapping results of Figure 5 along with a discussion of the temperature ensemble involved for this example can be found in [14]. Briefly summarized, as with the previous 13-atom Lennard-Jones example, Figure 5 denotes the results of a series of relaxation experiments. Here, however, 45 temperatures are used with the lowest 15 being involved in the heating/cooling process. The heating and cooling cycles consist of 1,200 smart Monte Carlo moves, each of one unit Lennard-Jones time duration. The cooling segment is taken as the portion of the cycle from moves 200 to 800 with the remainder being the heating portion. During the cooling portion of the cycle the 45 temperatures in the ensemble cover the range from (0.050–0.210) in temperature steps of 0.005, and from (0.210–0.330) in steps of 0.010, while during the heating portion of the cycle temperatures less than or equal to 0.150 are set equal to 0.150. The results shown in Figure 5 are obtained using 600 thermal cycles.

The 38-atom Lennard-Jones cluster has an interesting landscape. In particular, while the global and lowest-lying local minima are similar in energy, the minimum energy pathway that separates them involves appreciably higher energies and contains 13 separate barriers [24, Chapter 8.3]. As discussed in [14], the partial infinite swapping approach is appreciably more effective than conventional tempering approaches in providing a proper sampling of this complex potential energy landscape.

8. Appendix.

8.1. Proof of Theorem 4.1. Throughout the proof, we let $S = S_1 \times S_1$, where $S_1 \subset \mathbb{R}^d$ is convex and compact. Let $\mathcal{P}(S)$ denote the Polish space of all probability measures on S equipped with the topology of weak convergence. For any probability measure $\nu \in \mathcal{P}(S)$, define its mirror image $\nu^R \in \mathcal{P}(S)$ by requiring

$$\nu^R(A \times B) = \nu(B \times A)$$

for all Borel sets $A, B \subset \mathbb{R}^d$. Furthermore, as in the rest of the paper, a bold symbol $x \in S$ means $x = (x_1, x_2)$, where $x_1, x_2 \in \mathbb{R}^d$, and $x^R = (x_2, x_1)$. We also use the notation

$$\alpha(x, dy) \doteq \alpha_1(x_1, dy_1)\alpha_2(x_2, dy_2),$$

which is a probability transition kernel defined on S given S.

To prove the uniform LDP, it suffices to prove the equivalent uniform Laplace principle [3, Chapter 1]. To simplify the proof we have assumed that S is compact. This would be the case if, e.g., V is defined with periodic boundary conditions. The general case can be handled under (2.1) by using V as a Lyapunov function [3, section 8.2]. It will be convenient to split this into upper and lower bounds. We also consider just the (more complicated) case where $a_T \to \infty$ but $a_T < \infty$ for each T. Allowing $a_T = \infty$ requires a different notation to handle this special case but does not change the structure of the proof otherwise.

We will show for any bounded continuous function $F : \mathcal{P}(S) \to \mathbb{R}$ that

$$\lim_{T\to\infty} -\frac{1}{T} \log E\left[\exp\{-TF(\eta_T^{a_T})\}\right] = \inf_{\nu\in\mathcal{P}(S)} \left[F(\nu) + I^\infty(\nu)\right]. \tag{8.1}$$

By adding a constant to both sides we can assume $F \geq 0$, and we do so for the rest of this section.

The proof of the uniform LDP is based on the weak convergence approach. The proof is complicated by the multiscale aspect of the fast swapping process, as well as the fact that $\eta_T^{a_T}$ is a weighted empirical measure that involves this fast process.

8.2. Preliminary results.

8.2.1. A representation. We first state a stochastic control representation for the left-hand side of (8.1). As with the derivation of the infinite swapping process via weak convergence, it will be necessary to work with the (distributionally equivalent) temperature swapped processes for tightness to hold. In the representation, all random variables used to construct η_T^{aT} are replaced by random variables whose distribution is selected, and both the distributions and the random variables will be distinguished from their uncontrolled, original counterparts by an overbar. For this reason, while the continuous time process is denoted by $Y^a(t)$, we change notation and use $U^a(j)$ rather than $\bar{Y}^a(j)$ to denote the discrete time process.

We first construct the temperature swapped process. Let $\alpha(x, dy|0) = \alpha(x, dy)$ and $\alpha(x, dy|1) = \alpha(x^R, dy^R)$, and let $\{N_j^a,\ j = 0, 1, \ldots\}$ be i.i.d. geometric random variables with parameter $1/(1+a)$, i.e., geometric random variables with mean a. Then the random variables $\{U^a(j),\ j = 0, 1, \ldots\}$, $\{M_\ell^a(j),\ j = 0, 1, \ldots,\ \ell = 0, 1, \ldots, N_j^a\}$ are constructed recursively as follows. Given $M_0^a(j) = z$ and $U^a(j) = x$, $U^a(j+1)$ is distributed according to $\alpha(x, dy|z)$. The process $M_\ell^a(j)$, $\ell = 0, 1, \ldots, N_j^a$, is a Markov chain with states $\{0, 1\}$ and transition probabilities

$$\begin{array}{ll} p(0,0|x) = g(x), & p(0,1|x) = 1 - g(x), \\ p(1,0|x) = 1 - g(x^R), & p(1,0|x) = g(x^R). \end{array} \tag{8.2}$$

The initial value for the subsequent interval is given by $M_0^a(j+1) = M_{N_j^a}^a(j)$. Letting $\{\tau_{j,\ell}^a,\ i = 0,1,\ldots,\ \ell = 0,1,\ldots,N_j^a - 1\}$ be i.i.d. exponential random variables with mean $1/a$, the temperature swapped process in continuous time is then given by

$$Y^a(t) = U^a(j) \quad \text{for} \sum_{i=0}^{j-1}\sum_{\ell=0}^{N_j^a-1} \tau_{i,\ell}^a \le t < \sum_{i=0}^{j}\sum_{\ell=0}^{N_j^a-1} \tau_{i,\ell}^a$$

(with the convention that the sum from 0 to -1 is 0) and

$$Z^a(t) = M_\ell^a(j) \quad \text{for} \sum_{i=0}^{j-1}\sum_{k=0}^{\ell-1} \tau_{i,k}^a \le t < \sum_{i=0}^{j-1}\sum_{k=0}^{\ell} \tau_{i,k}^a,$$

and, last, the ordinary and weighted empirical measures are given by

$$\psi_T^a = \frac{1}{T}\int_0^T \delta_{Y^a(t)}dt \quad \text{and} \quad \eta_T^a = \frac{1}{T}\int_0^T \left[1_{\{Z^a(t)=0\}}\delta_{Y^a(t)} + 1_{\{Z^a(t)=1\}}\delta_{Y^a(t)^R}\right] dt.$$

Let σ^a denote the exponential distribution with mean $1/a$, and let β^a denote the geometric distribution with mean a. For the representation, all distributions (e.g., $\alpha(x, dy|z)$) can be perturbed from their original form, but such a perturbation pays a relative entropy cost. We distinguish the new distributions and random variables by using an overbar. Given $T \in (0,\infty)$, let R^a and K^a be the discrete time indices when the continuous time parameter reaches T, i.e.,

$$\sum_{i=0}^{R^a-2}\sum_{k=0}^{N_i^a-1} \tau_{i,k}^a + \sum_{k=0}^{K^a-1} \tau_{R^a-1,k}^a \le T < \sum_{i=0}^{R^a-2}\sum_{k=0}^{N_i^a-1} \tau_{i,k}^a + \sum_{k=0}^{K^a} \tau_{R^a-1,k}^a. \tag{8.3}$$

In this representation the barred quantities are constructed analogously to their unbarred counterparts. Thus, e.g., $\bar{R}^a$ and $\bar{N}_i^a$ are defined by (8.3) but with $\tau_{i,k}^a$ replaced by $\bar{\tau}_{i,k}^a$. Random variables corresponding to any given value of j are constructed in the order $\bar{U}^a(j+1), \bar{N}_j^a, \bar{M}_\ell^a(j), \bar{\tau}_{i,\ell}^a,\ \ell = 0,1,\ldots,\bar{N}_j^a$, and then j is updated to $j+1$. Barred measures, which are also allowed to depend on discrete time, are used to construct the corresponding barred random variables; e.g., $\bar{U}^a(j+1)$ is (conditionally) distributed according to $\bar{\alpha}_j(\bar{U}^a(j), \cdot|\bar{M}_0^a(j))$. The infimum is over all collections of measures $\{\bar{\alpha}_j, \bar{\beta}_j^a, \bar{p}_{j,\ell}, \bar{\sigma}_{j,\ell}^a\}$, and, although this is not denoted explicitly, any particular measure can depend on all previously constructed random variables. To simplify notation we let $\bar{N}_{\bar{R}^a}^a$ denote $\bar{K}^a$. We state the representation for $\{\eta_T^a\}$ and note that an analogous representation holds for $\{\psi_T^a\}$.

LEMMA 8.1. *Let $G : \mathcal{P}(S) \times \mathbb{N} \to \mathbb{R}$ be bounded from below and measurable. Then the representation*

$$-\frac{1}{T}\log E\left[\exp\{-TG(\eta_T^a, R^a)\}\right] = \inf E\left[G(\bar{\eta}_T^a, \bar{R}^a)\right.$$
$$+\frac{1}{T}\sum_{i=0}^{\bar{R}^a-1}\left[R\left(\bar{\alpha}_i(\bar{U}^a(i),\cdot|\bar{M}_0^a(i))\,\middle\|\,\alpha(\bar{U}^a(i),\cdot|\bar{M}_0^a(i))\right)+R\left(\bar{\beta}_i^a\,\middle\|\,\beta^a\right)\right]$$
$$\left.+\frac{1}{T}\sum_{i=0}^{\bar{R}^a-1}\sum_{k=0}^{\bar{N}_i^a-1}\left[R\left(\bar{p}_{i,k}(\bar{M}_k^a(i),\cdot|\bar{U}^a(i))\,\middle\|\,p(\bar{M}_k^a(i),\cdot|\bar{U}^a(i))\right)+R\left(\bar{\sigma}_{i,k}^a\,\middle\|\,\sigma^a\right)\right]\right]$$

is valid.

The proofs of such representations follow from the chain rule for relative entropy (see, e.g., [3, section B.2]). A novel feature of the representation here is that the total number of discrete time steps is random. However, this case can easily be reduced to the case with a fixed deterministic number of steps.

8.2.2. Rate for the ordinary empirical measure. We will frequently factor measures on product spaces in the proof. For a (deterministic) probability measure ν on a product space such as $S^1 \times S^2 \times S^3$, with each S^i a Polish space, we use notation such as $\nu_{1,2}$ to denote the marginal distribution on the first two components and notation such as $\nu_{1|3}$ to denote the conditional distribution on the first component given the third. When ν is a measurable random measure these can all be chosen so that they are also measurable.

We will make use of the rate function for the ordinary empirical measure. Let $\varphi(x, dy) = \alpha(x, dy|0)\rho_0(x) + \alpha(x, dy|1)\rho_1(x)$, where $\rho_0(x) = \rho(x)$ and $\rho_1(x) = \rho(x^R)$, and let $\bar{\mu}(dx) = [\pi(x) + \pi(x^R)]dx/2$ be its unique invariant probability distribution. If γ is absolutely continuous with respect to $\bar{\mu}$ with $\kappa(x) = [d\gamma/d\bar{\mu}](x)$, then set

$$K(\gamma) = 1 - \int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x, dy).$$

Note that K is convex. We then extend the definition to all of $\mathcal{P}(S)$ via lower semi-continuous regularization with respect to the weak topology. Thus if $\gamma_i \to \gamma$ in the weak topology and if each γ_i is absolutely continuous with respect to $\bar{\mu}$, then $\liminf_i K(\gamma_i) \geq K(\gamma)$, and we have equality for at least one such sequence. Note that since $\bar{\mu}$ is mutually absolutely continuous with respect to Lebesgue measure, this means that $K(\gamma) \leq 1$ for all $\gamma \in \mathcal{P}(S)$.

The following lemma will help relate weak limits of quantities in the representations to the rate function of the ordinary empirical measure.

LEMMA 8.2. *Let $\gamma \in \mathcal{P}(S)$ be absolutely continuous with respect to $\bar{\mu}$, and let $\kappa = [d\gamma/d\bar{\mu}]$. Assume $A \in (0, \infty)$, $\nu \in \mathcal{P}(S \times S)$ is such that $[\nu]_1 = [\nu]_2$, $R(\nu\|[\nu]_1 \otimes \varphi) < \infty$, r is such that $r[\nu]_1 = \kappa\bar{\mu}$, and $0 \leq -\int \log r(y)[\nu]_1(dy) < \infty$. Then*

$$K(\gamma) = 1 - \int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x, dy) \tag{8.4}$$
$$\leq AR(\nu\|[\nu]_1 \otimes \varphi) - A\int \log r(y)[\nu]_1(dy) + A\log A - A + 1.$$

Proof. Let C be the set where $\kappa(x) = 0$. Then $-\int \log r(y)[\nu]_1(dy) < \infty$ implies $r(x) > 0$ a.e. with respect to $[\nu]_1(dy)$, and we also have $\int r(y)[\nu]_1(dy) = 1$, so that

$r(y) < \infty$ a.s. with respect to $[\nu]_1(dy)$. It follows that $[\nu]_1(C) = 0$. Now suppose that

$$\int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x,dy) = 0.$$

Then $\kappa(x)\kappa(y) = 0$ a.s. with respect to Lebesgue measure, and so if $\kappa(x) \neq 0$, then $\varphi(x,\{y : \kappa(y) > 0\}) = 0$, or $\varphi(x,C) = 1$. Thus $\nu((S\backslash C) \times C) = 0$, while $[\nu]_1 \otimes \varphi((S\backslash C) \times C) = 1$, which implies $R(\nu\|[\nu]_1 \otimes \varphi) = \infty$, which is a contradiction. We conclude that

$$\int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x,dy) > 0.$$

Since also

$$\int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x,dy) \leq 1,$$

it follows that

$$-\log \int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x,dy) \in [0,\infty).$$

Since $R(\nu\|[\nu]_1 \otimes \varphi) < \infty$ and since $\bar{\mu}$ is the invariant distribution of φ, it follows that $R([\nu]_1\|\bar{\mu}) < \infty$ [3, Lemma 8.6.2]. This means that

$$\int \log \frac{\kappa(x)}{r(x)}[\nu]_1(dx) < \infty,$$

and since $-\int \log r(y)[\nu]_1(dy) < \infty$, it follows that $-\int \log \kappa(x)[\nu]_1(dx) > -\infty$. From $[\nu]_1 = [\nu]_2$ we conclude that

$$-\frac{1}{2}\int [\log \kappa(x) + \log \kappa(y)]\,\nu(dx,dy) > -\infty. \tag{8.5}$$

By relative entropy duality [3, Proposition 1.4.2]

$$\begin{aligned} &-\log \int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x,dy) \\ &\quad = -\log \int e^{\frac{1}{2}[\log \kappa(x) + \log \kappa(y)]}\bar{\mu}(dx)\varphi(x,dy) \\ &\quad \leq R(\nu\|\bar{\mu} \otimes \varphi) - \frac{1}{2}\int [\log \kappa(x) + \log \kappa(y)]\,\nu(dx,dy) \end{aligned}$$

is valid as long as the right-hand side is not of the form $\infty - \infty$, which is true by (8.5). The chain rule then gives

$$\begin{aligned}
&-\log \int e^{\frac{1}{2}[\log \kappa(x)+\log \kappa(y)]} \bar{\mu}(dx)\varphi(x,dy) \\
&\leq R(\nu\|\bar{\mu}\otimes\varphi) - \int \log\kappa(x)[\nu]_1(dx) \\
&= R(\nu\|[\nu]_1\otimes\varphi) + R([\nu]_1\|\bar{\mu}) - \int \log\kappa(x)[\nu]_1(dx) \\
&= R(\nu\|[\nu]_1\otimes\varphi) - \int \log r(x)[\nu]_1(dx),
\end{aligned}$$

and thus

$$-\int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x,dy) \leq -e^{-R(\nu\|[\nu]_1\otimes\varphi)+\int \log r(x)[\nu]_1(dx)}.$$

Then (8.4) follows from the fact that if $a \in \mathbb{R}$ and $b \in (0,\infty)$, then $-e^{-a} \leq ab + b\log b - b$ by taking $a = R(\nu\|[\nu]_1\otimes\varphi) - \int \log r(x)[\nu]_1(dx)$ and $b = A$. □

8.2.3. Decomposition of the exponential distribution. In the construction of η_T^a we used independent exponential random variables $\tau_{i,k}^a$ and geometric random variables N_i^a and the fact that for each i

$$\sum_{\ell=0}^{N_i^a-1} \tau_{i,\ell}^a$$

is exponential with mean one. This decomposition corresponds to a relationship in relative entropies, which we now state.

LEMMA 8.3. *For $a \in (0,\infty)$, let $\bar{N}^a$ be distributed according to a random probability measure $\bar{\beta}^a$ on $\{0,1,\ldots\}$. Given $\bar{N}^a = \ell$, for $k \in \{0,1,\ldots l\}$ let $\bar{\tau}_k^a(\ell)$ be distributed according to a random probability measure $\bar{\sigma}_k^a(\ell)$ on $[0,\infty)$. Let $\bar{\sigma}$ be the distribution of the random variable*

$$\bar{\tau} = \sum_{k=0}^{\bar{N}^a-1} \bar{\tau}_k^a(\bar{N}^a).$$

Then

$$E\left[R\left(\bar{\beta}^a\|\beta^a\right) + \sum_{k=0}^{\bar{N}^a-1} R\left(\bar{\sigma}_k^a(\bar{N}^a)\|\sigma^a\right)\right] \geq E\left[R\left(\bar{\sigma}\|\sigma^1\right)\right].$$

Proof. Define a measure μ on the space $\mathbb{N}_+ \times \prod_{i=0}^{\infty}[0,\infty)$ as follows. For any $\ell \in \mathbb{N}_+$ and any sequence $A_0, A_1, \ldots$ of Borel measurable subsets of $[0,\infty)$,

$$\mu\left(\{l\}\times A_0\times A_1\times\cdots\right) = \beta^a(l)\prod_{k=0}^{\infty}\sigma^a(A_k),$$

and similarly define the measure $\bar{\mu}$ by

$$\bar{\mu}\left(\{l\} \times A_0 \times A_1 \times \cdots\right) = \bar{\beta}^a(l) \left(\prod_{k=0}^{\ell-1} \bar{\sigma}_k^a(\ell)(A_k)\right) \left(\prod_{k=\ell}^{\infty} \sigma^a(A_k)\right).$$

Then by the chain rule of relative entropy

$$E\left[R\left(\bar{\mu} \,\|\mu\right)\right] = E\left[R\left(\bar{\beta}^a \,\|\beta^a\right) + \sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell-1} R\left(\bar{\sigma}_k^a(\ell) \,\|\sigma^a\right) \bar{\beta}^a(\ell)\right].$$

Since

$$\begin{aligned} E\left[\sum_{k=0}^{\bar{N}^a-1} R\left(\bar{\sigma}_k^a(\bar{N}^a) \,\|\sigma^a\right)\right] &= E\left[E\left[\sum_{k=0}^{\bar{N}^a-1} R\left(\bar{\sigma}_k^a(\bar{N}^a) \,\|\sigma^a\right) \middle| \bar{N}^a\right]\right] \\ &= E\left[\sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell-1} R\left(\bar{\sigma}_k^a(\ell) \,\|\sigma^a\right) \bar{\beta}^a(\ell)\right], \end{aligned}$$

it follows that

$$E\left[R\left(\bar{\mu} \,\|\mu\right)\right] = E\left[R\left(\bar{\beta}^a \,\|\beta^a\right) + \sum_{k=0}^{\bar{N}^a-1} R\left(\bar{\sigma}_k^a(\bar{N}^a) \,\|\sigma^a\right)\right].$$

Observe that $\bar{\tau}$ can be written as a measurable mapping on $\mathbb{N}_+ \times \prod_{i=0}^{\infty}[0,\infty)$ and that $\bar{\sigma}$ and σ^1 are the distributions induced on $[0,\infty)$ under that map by $\bar{\mu}$ and μ, respectively. Since relative entropy can only decrease under such a mapping, it follows that $ER\left(\bar{\mu} \,\|\mu\right) \geq ER\left(\bar{\sigma} \,\|\sigma^1\right)$. □

8.3. Lower bound. The proof of the lower bound will be partitioned into three cases according to a parameter $C \in (1,\infty)$. After the three cases have been argued, the proof of the lower bound will be completed. The first two cases are very simple, and they give estimates when R^a/T is small (i.e., unusually few exponential clocks with mean 1 are needed before time T is reached) or when R^a/T is large (i.e., an unusually large number of such clocks are needed to reach time T). The processes $\{U^a(j)\}$ and $\{M_\ell^a(j)\}$ play no role in these estimates, and the required estimates follow from Chebyshev's inequality as it is used in the proof of Crámer's theorem. We will need the function $h(b) = -\log b + b - 1$, $b \in [0,\infty)$, which satisfies $\inf\left\{R\left(\gamma \,\|\sigma^1\right) : \int u\gamma(du) = b\right\} = h(b)$.

8.3.1. The case $R^a/T \leq 1/C$. Let $F : \mathcal{P}(S) \to \mathbb{R}$ be nonnegative and continuous. Then

$$\begin{aligned} &-\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^a) - \infty 1_{\{[0,1/C]\}^c}(R^a/T)\}\right] \\ &\quad = -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^a)\}\, 1_{\{[0,1/C]\}}(R^a/T)\right] \\ &\quad \geq -\frac{1}{T}\log P\left\{\sum_{i=0}^{\lfloor T/C\rfloor+1} \tau_i^1 \geq T\right\}. \end{aligned}$$

By Chebyshev's inequality, for $\alpha \in (0,1)$,

$$\begin{aligned} P\left\{\sum_{i=0}^{\lfloor T/C\rfloor+1} \tau_i^1 \geq T\right\} &= P\left\{e^{\alpha\sum_{i=0}^{\lfloor T/C\rfloor+1}\tau_i^1} \geq e^{\alpha T}\right\} \\ &\leq \exp\left\{(\lfloor T/C\rfloor + 2)\left(\log\frac{1}{1-\alpha} - \alpha C\right)\right\}. \end{aligned}$$

Optimizing this inequality over $\alpha \in (0,1)$ gives

$$\begin{aligned} &\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^a) - \infty 1_{\{[0,1/C]\}^c}(R^a/T)\}\right] \\ &\quad \geq \liminf_{T\to\infty}\frac{1}{T}\left(\lfloor T/C\rfloor + 2\right)h(C) \\ &\quad = \frac{h(C)}{C}. \end{aligned}$$

Note that $h(C)/C \to 1$ as $C \to \infty$.

8.3.2. The case $R^a/T \geq C$. With F as in the last section, an analogous argument gives

$$\begin{aligned} &\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^a) - \infty 1_{\{[C,\infty)\}^c}(R^a/T)\}\right] \\ &\quad \geq \liminf_{T\to\infty}\frac{1}{T}\left(\lfloor T(C-1)\rfloor + 2\right)h\left(\frac{1}{C-1}\right) \\ &\quad = (C-1)\,h\left(\frac{1}{C-1}\right). \end{aligned}$$

Note that $(C-1)\,h(1/(C-1)) \to \infty$ as $C \to \infty$.

8.3.3. The case $1/C \leq R^a/T \leq C$. To analyze this case it will be sufficient to consider any deterministic sequence $\{r^a\}$ such that $1/C \leq r^a/T \leq C$, and such that $r^a/T \to A$ as $T \to \infty$, and obtain lower bounds on

$$\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^a) - \infty 1_{\{r^a/T\}^c}(R^a/T)\}\right].$$

Since I^∞ is convex, to prove the lower bound for (8.1) we can assume F is convex and lower semicontinuous [3, Theorem 1.2.1]. The representation from Lemma 8.1 will be applied. We first note that the representation will include a term of the form $\infty 1_{\{r^a/T\}^c}(\bar{R}^a/T)$ on the right-hand side. We can remove this term if we restrict the infimum to controls for which $\bar{R}^a = r^a$ w.p.1, and we do so for notational convenience. The representation thus becomes

(8.6)

$$\begin{aligned}
&-\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^a) - \infty 1_{\{r^a\}^c}(R^a)\}\right] \\
&= \inf E\left[F(\bar{\eta}_T^a) + \frac{1}{T}\sum_{i=0}^{r^a-1}\left[R\left(\bar{\alpha}_i(\bar{U}^a(i),\cdot|\bar{M}_0^a(i))\,\middle\|\,\alpha(\bar{U}^a(i),\cdot|\bar{M}_0^a(i))\right)\right.\right. \\
&\qquad + R\left(\bar{\beta}_i^a\,\|\beta^a\right)\Big] + \frac{1}{T}\sum_{i=0}^{r^a-1}\sum_{k=0}^{\bar{N}_i^a-1}\left[R\left(\bar{p}_{i,k}(\bar{M}_k^a(i),\cdot|\bar{U}^a(i))\,\middle\|\,p(\bar{M}_k^a(i),\cdot|\bar{U}^a(i))\right)\right. \\
&\qquad \left.\left. + R\left(\bar{\sigma}_{i,k}^a\,\|\sigma^a\right)\right]\right].
\end{aligned}$$

There are four relative entropy sums in (8.6). Since F is bounded from below, the lower bound holds vacuously unless each such term is uniformly bounded.

We first show that the empirical distribution on $\bar{N}_i^a$ converges to δ_∞ by using a martingale argument. For $C_1, C_2 \subset [0,\infty)$, let

$$\xi^T(C_1\times C_2) \doteq \frac{1}{r^a}\sum_{i=0}^{r^a-1}\delta_{\bar{N}_i^a}(C_1)\bar{\beta}_i^a(C_2).$$

Consider the one-point compactification of $[0,\infty)$, which we identify with $[0,\infty]$, and left $f:[0,\infty]\to\mathbb{R}$ be bounded and continuous. Then for any $\epsilon > 0$,

$$
\begin{aligned}
(8.7)\qquad & P\left\{\left|\int [f(z_1)-f(z_2)]\xi^T(dz_1\times dz_2)\right|\geq\varepsilon\right\}\\
&= P\left\{\left|\frac{1}{r^a}\sum_{i=0}^{r^a-1}\left(f(\bar{N}_i^a)-\int f(n)\bar{\beta}_i^a(dn)\right)\right|\geq\varepsilon\right\}\\
&\leq \frac{1}{\varepsilon^2}E\left[\left|\frac{1}{r^a}\sum_{i=0}^{r^a-1}\left(f(\bar{N}_i^a)-\int f(n)\bar{\beta}_i^a(dn)\right)\right|^2\right]\\
&\leq \frac{\|f\|_\infty^2}{\varepsilon^2 r^a}\\
&\to 0,
\end{aligned}
$$

and thus any weak limit of ξ^T has identical marginals. Next note that by Jensen's inequality and since $r^a/T\to A\in(0,\infty)$, the uniform bound on the second relative entropy sum implies that $ER\left([\xi^T]_2\,\|\beta^a\right)$ is uniformly bounded. Using that $\inf\{R(\gamma\,\|\beta_a):\int u\gamma(du)=b\}=b\log\frac{b}{a}+(1+b)\log\frac{1+a}{1+b}$, it follows that the weak limit of $[\xi^T]_2=\delta_\infty$ w.p.1.

Next we consider the asymptotic properties of the collection $\{\bar{M}_\ell^a(i)\}$ under boundedness of the third relative entropy sum. We use that the empirical measure of the $\{\bar{N}_i^a\}$ tends to δ_∞. Since $p(\cdot,\cdot|\bar{U}^a(i))$ is the transition function of a finite state Markov chain, this means that asymptotically the $\bar{M}_k^a(i)$, $k=0,\ldots,\bar{N}_i^a$, are samples from the transition probability (8.2) with $x=\bar{U}^a(i)$, and in particular that $\bar{M}_0^a(i+1)$ is asymptotically conditionally independent with distribution $(\rho(x),1-\rho(x))$. Indeed, a martingale argument similar to (8.7) shows that if

$$
\mu^T(C_1\times C_2)\doteq\frac{1}{r^a}\sum_{i=0}^{r^a-1}\delta_{\bar{U}^a(i)}(C_1)\delta_{\bar{M}_0^a(i)}(C_2)
$$

and if μ^∞ is the weak limit of any convergent subsequence (which must exist by compactness), then

$$
(8.8)\qquad \left([\mu^\infty]_{2|1}(\{0\}\,|y),[\mu^\infty]_{2|1}(\{1\}\,|y)\right)=(\rho(y),1-\rho(y))\ \ [\mu^\infty]_1\text{-a.s.},
$$

and the same is true for the empirical measure of $\left\{\bar{M}_k^a(i),\,k=0,\ldots,\bar{N}_i^a\right\}$.

We next remove the third relative entropy sum from the representation (8.6) and obtain a lower bound for the right-hand side using Lemma 8.3. Let $\hat{\sigma}_i^a$ denote the distribution of the random variable

$$
\hat{\tau}_i^a=\sum_{\ell=0}^{\bar{N}_i^a-1}\bar{\tau}_{i,\ell}^a.
$$

Then we have the lower bound

$$
\begin{aligned}
(8.9)\qquad & -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^a)-\infty 1_{\{r^a\}^c}(R^a)\}\right]\\
&\geq \inf\left[F(E\bar{\eta}_T^a)+\frac{1}{T}E\sum_{i=0}^{r^a-1}\left[R\left(\bar{\alpha}_i(\bar{U}^a(i),\cdot|\bar{M}_0^a(i))\,\|\alpha(\bar{U}^a(i),\cdot|\bar{M}_0^a(i))\right)\right.\right.\\
&\qquad\left.\left.+R\left(\hat{\sigma}_i^a\,\|\sigma^1\right)\right]\right].
\end{aligned}
$$

To study the lower bound of (8.9) we introduce the measures

$$\kappa^T(C_1 \times C_2 \times C_3) \doteq \frac{1}{r^a} \sum_{i=0}^{r^a-1} \delta_{\bar{U}^a(i)}(C_1)\bar{\alpha}_i(\bar{U}^a(i), C_2|\bar{M}_0^a(i))\delta_{\bar{M}_0^a(i)}(C_3),$$

$$\xi^T(C_1 \times C_2) \doteq \frac{1}{T} \sum_{i=0}^{r^a-1} \delta_{\bar{U}^a(i)}(C_1)\delta_{(\hat{m}_i^a)^{-1}}(C_2)\hat{m}_i^a,$$

where $\hat{m}_i^a$ is the conditional mean of $\hat{\tau}_i^a$. The restriction on the control measures implies

$$\sum_{i=0}^{r^a-2} \hat{\tau}_i^a \leq T < \sum_{i=0}^{r^a-1} \hat{\tau}_i^a,$$

and since function h is increasing on $(1, \infty)$, we can assume without loss of generality that $\hat{m}_{r^a-1}^a \leq 1$.

We introduce the function $\ell : \mathbb{R}_+ \to \mathbb{R}_+$ given by $\ell(b) = b\log b - b + 1$. Note that $h(b) = b\ell(1/b)$. The sequence $\{\kappa^T\}$ is tight because S and $\{0,1\}$ are compact. Using the fact that $\hat{\sigma}_i^a$ selects the conditional distribution of $\hat{\tau}_i^a$ and $\inf\{R(\gamma \| \sigma^1) : \int u\gamma(du) = b\} = h(b)$, we have

$$\begin{aligned} E\left[\frac{1}{T}\sum_{i=0}^{r^a-1} R(\hat{\sigma}_i^a \| \sigma^1)\right] &\geq E\left[\frac{1}{T}\sum_{i=0}^{r^a-1} h(\hat{m}_i^a)\right] \\ &= E\left[\frac{1}{T}\sum_{i=0}^{r^a-1} \hat{m}_i^a \ell([\hat{m}_i^a]^{-1})\right] \\ &= E\left[\int \ell(z)\xi^T(du \times dz)\right]. \end{aligned}$$

Hence the uniform bound on the relative entropy sum gives that $\{\xi^T\}$ is tight. Let $f : S \to \mathbb{R}$ be bounded and continuous. Since $\bar{\alpha}_i(\bar{U}^a(i), \cdot|\bar{M}_0^a(i))$ selects the conditional distribution of $\bar{U}^a(i+1)$, for $\varepsilon > 0$,

$$
\begin{aligned}
&P\left\{\left|\sum_{z=1}^{2}\int[f(y_1)-f(y_2)]\kappa^T(dy_1\times dy_2\times dz)\right|\geq\varepsilon\right\}\\
&=P\left\{\left|\frac{1}{r^a}\sum_{i=0}^{r^a-1}\left(f(\bar{U}^a(i+1))-\int f(y)\bar{\alpha}_i(\bar{U}^a(i),dy|\bar{M}_0^a(i))\right)\right|\geq\varepsilon\right\}\\
&\leq\frac{1}{\varepsilon^2}E\left[\left|\frac{1}{r^a}\sum_{i=0}^{r^a-1}\left(f(\bar{U}^a(i+1))-\int f(y)\bar{\alpha}_i(\bar{U}^a(i),dy|\bar{M}_0^a(i))\right)\right|^2\right]\\
&\leq\frac{\|f\|_\infty^2}{\varepsilon^2 r^a}\\
&\to 0.
\end{aligned}
$$

We find that

$$[\kappa^\infty]_1=[\kappa^\infty]_2 \text{ w.p.1.} \tag{8.10}$$

Using Fatou's lemma and the definition of φ, we get the lower bound $AR([\kappa^\infty]_{1,2}\|[\kappa^\infty]_1\otimes\varphi)$ on the weak limit of the corresponding relative entropies (the second sum in (8.9)).

With regard to ξ^∞, comparing the form of $\left[\xi^T\right]_1$ and $\bar{\psi}_T^a$ gives

$$E\left[\xi^\infty\right]_1=E\bar{\psi}_\infty.$$

Observe that

$$\int z\xi^T(dx\times dz)=\frac{r^a}{T}\left[\kappa^T\right]_1(dx).$$

Because of the superlinearity of ℓ, we have uniform integrability, and thus passing to the limit gives

$$\int z[\xi^\infty]_1(dx)[\xi^\infty]_{2|1}(dz|x)=A[\kappa^\infty]_1(dx).$$

Hence with the definition $b(x)=\int z[\xi^\infty]_{2|1}(dz|x)$, $[d[\kappa^\infty]_1/d[\xi^\infty]_1](x)=b(x)/A$. Using Fatou's lemma, Jensen's inequality, and the weak convergence, we get the following lower bound on the corresponding relative entropies (the first sum in (8.9)):

$$\begin{aligned}
&\liminf_{T\to\infty} \int \ell(z)\xi^T(du\times dz)\\
&\qquad \geq \int \ell(z)\xi^\infty(du\times dz)\\
&\qquad \geq \int \ell\left(\int z[\xi^\infty]_{2|1}(dz|x)\right)[\xi^\infty]_1(dx)\\
&\qquad = \int \ell\left(b(x)\right)[\xi^\infty]_1(dx)\\
&\qquad = A\int h(1/b(x))[\kappa^\infty]_1(dx).
\end{aligned}$$

Let $r(y) = A/b(y)$. Then we can write the combined lower bound on the relative entropies as

$$\begin{aligned}
&AER([\kappa^\infty]_{1,2}\|[\kappa^\infty]_1\otimes\varphi) + AE\int h(1/b(x))[\kappa^\infty]_1(dx)\\
&= AER([\kappa^\infty]_{1,2}\|[\kappa^\infty]_1\otimes\varphi) - AE\int \log r(y)[\kappa^\infty]_1(dy) + A\log A - A + 1. \qquad (8.11)
\end{aligned}$$

We next consider $\bar{\psi}^a_T$ and $\bar{\eta}^a_T$. Using the limiting properties of the $\bar{M}^a_k(i)$, we have

$$\begin{aligned}
\bar{\eta}^a_T(C) &= \frac{1}{T}\int_0^T \left[1_{\{\bar{Z}^a(t)=0\}}\delta_{\bar{Y}^a(t)}(C) + 1_{\{\bar{Z}^a(t)=1\}}\delta_{\bar{Y}^a(t)^R}(C)\right]dt\\
&\to \int \left[\rho(y)\delta_y(C) + (1-\rho(y))\delta_{y^R}(C)\right][\xi^\infty]_1\\
&= \bar{\eta}_\infty(C)
\end{aligned}$$

and

$$\bar{\psi}^a_T(C) = \frac{1}{T}\int_0^T \delta_{\bar{Y}^a(t)}(C)dt \to \int_C [\xi^\infty]_1 = \bar{\psi}_\infty(C).$$

Note that this implies the relation

$$\bar{\eta}_\infty(C) = \int_C \left[\rho(y)\bar{\psi}_\infty(dy) + \rho(y^R)\bar{\psi}_\infty(dy^R)\right]. \qquad (8.12)$$

Finally we consider the weighted empirical measure. By (8.11), Lemma 8.2, and Jensen's inequality we have the lower bound $\left[F(E\bar{\eta}_\infty)+K(E\bar{\psi}_\infty)\right]$ for the limit inferior of the right-hand side of (8.6), where $\bar{\eta}_\infty$ and $\bar{\psi}_\infty$ are related by (8.12). Thus we need only show that

$$K(E\bar{\psi}_\infty) \geq I^0(E\bar{\eta}_\infty) = I^\infty(E\bar{\eta}_\infty).$$

The equality follows from the definition of I^∞ and (8.12). Let $a(y) = [dE\bar{\psi}_\infty/d\bar{\mu}](y)$. Then

$$K(E\bar{\psi}_\infty) = 1 - \int \sqrt{a(x)a(y)}\bar{\mu}(dx)\varphi(x,dy)$$

and

$$I^0(E\bar{\eta}_\infty) = 1 - \int \frac{1}{2}\sqrt{(a(x)+a(x^R))\,(a(y)+a(y^R))}\mu(dx)\alpha(x,dy).$$

Using that $\rho(x) = 1-\rho(x^R)$ and symmetry,

$$\begin{aligned}
&\int \sqrt{a(x)a(y)}\frac{1}{2}\left[\mu(dx)+\mu(dx^R)\right]\left(\rho(x)\alpha(x,dy)+\rho(x^R)\alpha(x^R,dy^R)\right)\\
&=\frac{1}{2}\int\left(\sqrt{a(x)a(y)}+\sqrt{a(x^R)a(y^R)}\right)\mu(dx)\left(\rho(x)\alpha(x,dy)+\rho(x^R)\alpha(x^R,dy^R)\right)\\
&=\frac{1}{2}\int\left(\sqrt{a(x)a(y)}+\sqrt{a(x^R)a(y^R)}\right)\mu(dx)\alpha(x,dy).
\end{aligned}$$

We now use

$$\sqrt{a(x)a(y)}+\sqrt{a(x^R)a(y^R)} \le \sqrt{(a(x)+a(x^R))\,(a(y)+a(y^R))}$$

to obtain $K(E\bar{\psi}_\infty) \ge I^0(E\bar{\eta}_\infty)$. (We note for later use that given $\bar{\eta}$ that is absolutely continuous with respect to Lebesgue measure and such that $I^\infty(\bar{\eta}) < \infty$, $K(\bar{\psi}) = I^0(\bar{\eta})$ can be shown for a $\bar{\psi}$ that maps to $\bar{\eta}$ by taking $\bar{\psi} = [\bar{\eta}+\bar{\eta}^R]/2$.)

8.3.4. Combining the cases. In the last subsection we showed that for any sequence r^a such that $r^a/T \to A \in [1/C, C]$,

$$\begin{aligned}
\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^a) - \infty 1_{\{r^a/T\}^c}(R^a/T)\}\right] &\ge \left[F(E\bar{\eta}_\infty) + I^\infty(E\bar{\eta}_\infty)\right]\\
&\ge \inf_{\nu\in\mathcal{P}(S)}\left[F(\nu)+I^\infty(\nu)\right],
\end{aligned}$$

and an argument by contradiction shows that the bound is uniform in A. Thus

$$\begin{aligned}
&\liminf_{T\to\infty} -\frac{1}{T}\log\left\{\sum_{r^a=\lceil\frac{T}{C}\rceil}^{\lfloor TC\rfloor} E\left[\exp\{-TF(\eta_T^a) - \infty 1_{\{r^a/T\}^c}(R^a/T)\}\right]\right\}\\
&\quad\ge \liminf_{T\to\infty} -\frac{1}{T}\log\left\{TC\cdot\bigvee_{r^a=\lceil\frac{T}{C}\rceil}^{\lfloor TC\rfloor} E\left[\exp\{-TF(\eta_T^a) - \infty 1_{\{r^a/T\}^c}(R^a/T)\}\right]\right\}\\
&\quad\ge \inf_{\nu\in\mathcal{P}(S)}\left[F(\nu)+I^\infty(\nu)\right].
\end{aligned}$$

We now partition $E\left[\exp\{-TF(\eta_T^a)\}\right]$ according to the various cases to obtain the overall lower bound

$$\begin{aligned}
&\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^{a_T})\}\right]\\
&\qquad\ge \min\left\{\inf_{\nu\in\mathcal{P}(S)}\left[F(\nu)+I^\infty(\nu)\right], \frac{h(C)}{C}, (C-1)\,h\left(\frac{1}{C-1}\right)\right\}.
\end{aligned}$$

Letting $C \to \infty$ and using the fact that $I^\infty \leq 1$, we have the desired lower bound

$$\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^{a_T})\}\right] \geq \inf_{\nu\in\mathcal{P}(S)}\left[F(\nu)+I^\infty(\nu)\right].$$

8.4. **Upper bound.** The proof of the reverse inequality

$$\limsup_{T\to\infty} -\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^{a_T})\}\right] \leq \inf_{\nu\in\mathcal{P}(S)}\left[F(\nu)+I^\infty(\nu)\right] \tag{8.13}$$

is simpler. Let bounded and continuous F be given. Given $\varepsilon > 0$, we can find ν that is absolutely continuous with respect to Lebesgue measure and for which

$$\left[F(\nu)+I^\infty(\nu)\right] \leq \inf_{\nu\in\mathcal{P}(S)}\left[F(\nu)+I^\infty(\nu)\right]+\varepsilon.$$

Next we use that I^∞ is convex and $I^\infty(\mu) = 0$ to find $\tau > 0$ such that

$$\left[F(\nu^\tau)+I^\infty(\nu^\tau)\right] \leq \left[F(\nu)+I^\infty(\nu)\right]+\varepsilon$$

when $\nu^\tau = \tau\mu + (1-\tau)\nu$. Note that if $\theta(x) = [d\nu/d\mu](x)$ and $\theta^\tau(x) = [d\nu^\tau/d\mu](x)$, then $\theta^\tau(x) \geq \tau > 0$ and so $I^\infty(\nu^\tau) < 1$.

We will construct a control to use in the representation such that $\bar{\eta}_T^{a_T}$ will converge w.p.1 to ν^τ and RE^T will converge w.p.1 to $I^\infty(\nu^\tau)$. These convergences will follow from the ergodic theorem, and the bound $\theta^\tau(x) \geq \tau$ will be used to argue that the ergodicity of the original process is inherited by the controlled process.

We now proceed to the construction. Let $\varsigma(dx) = [\nu^\tau(dx)+\nu^\tau(dx^R)]/2$ so that $I^\infty(\nu^\tau) = K(\varsigma)$. Let $\kappa(x) = [d\varsigma/d\bar{\mu}](x) \geq \tau/2$. Then

$$K(\varsigma) = 1 - \int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x,dy).$$

Since κ is uniformly bounded from below, we have the dual relationship

$$\begin{aligned}
&-\log\int \sqrt{\kappa(x)\kappa(y)}\bar{\mu}(dx)\varphi(x,dy)\\
&\quad = -\log\int e^{\frac{1}{2}[\log\kappa(x)+\log\kappa(y)]}\bar{\mu}(dx)\varphi(x,dy)\\
&\quad = R(\eta\|\bar{\mu}\otimes\varphi) - \frac{1}{2}\int\left[\log\kappa(x)+\log\kappa(y)\right]\nu(dx,dy),
\end{aligned}$$

where

$$\eta(C) = \frac{\int_C e^{\frac{1}{2}[\log\kappa(x)+\log\kappa(y)]}\bar{\mu}(dx)\varphi(x,dy)}{\int_{S^2} e^{\frac{1}{2}[\log\kappa(x)+\log\kappa(y)]}\bar{\mu}(dx)\varphi(x,dy)}.$$

Since $\sqrt{\kappa(x)\kappa(y)}$ is symmetric in x and y, automatically $[\eta]_1 = [\eta]_2$, and using the bound $\kappa(x) \geq \tau/2$ we can factor $\eta(dx \times dy) = [\eta]_1(dx)\bar{\varphi}(x, dy)$, where $\bar{\varphi}$ is the transition kernel of a uniformly ergodic Markov chain. Let

$$\bar{\alpha}(x, dy|0) = \bar{\alpha}(x, dy|1) = \bar{\varphi}(x, dy).$$

Notice that from the definition of $\bar{\varphi}$, $\bar{\varphi}(x, dy) = \bar{\varphi}(x^R, dy^R)$, and hence $\bar{\alpha}$ has the property that

$$\bar{\alpha}(x^R, dy^R|0) = \bar{\alpha}(x, dy|1).$$

These will be the transition kernels used to construct the $\bar{U}^a(i)$.

Now of course the invariant distribution of $\bar{\varphi}$ is $[\eta]_1$ and not the desired distribution ς. Let $r(x) = [d\varsigma/d[\eta]_1](x)$. Then $r(x)$ identifies the way in which the distribution of the random variables $\bar{N}_j^a$ should be modified so that in the continuous time the empirical measure $[\eta]_1$ is reshaped into ς. Choose A so that equality holds in (8.4), and set $b(x) = Ar(x)$. Then $\bar{\beta}_i^a$ is chosen to be $\beta^{ab(x)}$. Consistent with the analysis of the upper bound, we do not perturb the distribution of the other variables so that $\bar{p}_{i,k}(\cdot, \cdot|\cdot) = p(\cdot, \cdot|\cdot)$ and $\bar{\sigma}_{j,\ell}^a = \sigma^a$. We then construct the controlled processes using these measures in exact analogy with the construction of the original process.

With this choice and Lemma 8.1 we obtain the bound

$$\begin{aligned}
&-\frac{1}{T}\log E\left[\exp\{-TF(\eta_T^{a\tau})\}\right] \\
&\quad\leq E\left[F(\bar{\eta}_T^a) + \frac{1}{T}\sum_{i=0}^{\bar{R}^a-1}\left[R\left(\bar{\alpha}(\bar{U}^a(i), \cdot|\bar{M}_0^a(i))\,\middle\|\,\alpha(\bar{U}^a(i), \cdot|\bar{M}_0^a(i))\right)\right.\right. \\
&\qquad\left.\left. + R\left(\beta^{ab(\bar{U}^a(i))}\,\middle\|\,\beta^a\right)\right]\right].
\end{aligned}$$

Now apply the ergodic theorem, and use that w.p.1 $A^T = \bar{R}^a/T \to 1/\int b(x)[\eta]_1(dx)$ $= A$ and also that asymptotically the conditional distribution of $\bar{M}_0^a(i)$ is given by $(\rho_0(\bar{U}^a(i)), \rho_1(\bar{U}^a(i)))$. Then the right-hand side of the last display converges until we have the limit

$$\begin{aligned}
&F(\nu^\tau) + \frac{1}{\int b(x)[\eta]_1(dx)}\int\left[\sum_{z=1}^{2} R\left(\bar{\alpha}(x, dy|z)\,\|\alpha(x, dy|z)\right)\rho_z(x)\right] r(x)[\eta]_1(dx) \\
&\quad + \frac{1}{\int b(x)[\eta]_1(dx)}\int\left(-\log b(x) + b(x) - 1\right) b(x)[\eta]_1(dx) \\
&= F(\nu^\tau) + A\int R\left(\bar{\varphi}(x, dy)\,\|\varphi(x, dy)\right)\varsigma(dx) - A\int \log r(x)\varsigma(dx) + A\log A - A + 1 \\
&= F(\nu^\tau) + K(\varsigma) \\
&= F(\nu^\tau) + I^\infty(\nu^\tau).
\end{aligned}$$

This completes the proof of (8.13) and also the proof of Theorem 4.1.

[0220] The foregoing has outlined some of the more pertinent features of the subject matter. These features should be construed to be merely illustrative. Many other beneficial results can be attained by applying the disclosed subject matter in a different manner or by modifying the subject matter as will be described. For example, density or pressure may be used as a control variable in the place of temperature.

CLAIMS:

1. A computer system for simulating a physical system using a control variable, comprising:

a grouping module for dividing a control variable range into a first set of groups and a second set of groups, the control variable range extending from a lowest value, V_{lowest}, to a highest value, $V_{highest}$, each group in the first set including a discrete set of control values that extend over only a portion of the range from V_{lowest} to $V_{highest}$, each group in the second set including a discrete set of control values that extend over only a portion of the range from V_{lowest} to $V_{highest}$, each group in the first set being distinct from each group in the second set, the control values in the first set extending from V_{lowest} to $V_{highest}$, the control values in the second set extending from V_{lowest} to $V_{highest}$;

an analysis module for computing data representative of the physical system over the set of control values in each group in the first and second set of groups, the analysis module providing simultaneous exchange of information between all of the control values within a group; and

a swapping module for communicating data such that:

the analysis module can use data computed using a group in the first set for computing data representative of the physical system for a group of control values in the second set and,

the analysis module can use data computed using a group in the second set for computing data representative of the physical system for a group of control values in the first set.

2. The computer system of claim 1, wherein the control values comprise one of density, pressure, volume and temperature.

3. The computer system of claim 1, wherein the physical system is a nanoscale structure.

4. A computer system for simulating a physical system using a control variable, comprising:

a first analysis module for computing data representative of the physical system over a range of a control variable, the range extending from a lowest value V_{lowest}, to a highest value, $V_{highest}$, the first analysis module individually computing data representative of the physical system within each of a first set of groups, each group in the first set including a

discrete set of values of the control variable, the discrete set in each group of the first set of groups extending over only a portion of the range from V_{lowest} to $V_{highest}$, the first analysis module providing simultaneous exchange of information between all values of the control variable within a group of the first set of groups;

a second analysis module for computing data representative of the physical system over the range of the control variable extending from V_{lowest} to $V_{highest}$, the second analysis module individually computing data representative of the physical system within each of a second set of groups, each group in the second set including a discrete set of values of the control variable, the discrete set in each group of the second set of groups extending over only a portion of the range from V_{lowest} to $V_{highest}$, the second analysis module providing simultaneous exchange of information between all values of the control variable within a group of the second set of groups, each group in the second set of groups being distinct from each group in the first set of groups; and

a swapping module for communicating data such that:

the second analysis module can receive data computed by the first analysis module for a group in the first set of groups, the second analysis module using that data received from the first analysis module for computing data representative of the physical system for a group in the second set of groups.

5. A computer system according to claim 4, the first analysis module and the second analysis module being formed from a common module of computer software.

6. A computer system according to claim 4, the swapping module also communicating data such that the first analysis module can receive data computed by the second analysis module for a group in the second set of groups, the first analysis module using that data received from the second analysis module for computing data representative of the physical system for a group in the first set of groups.

7. A computer system according to claim 4, further comprising a third analysis module for computing data representative of the physical system over the range of the control variable extending from V_{lowest} to $V_{highest}$, the third analysis module individually computing data representative of the physical system within each of a third set of groups, each group in the third set including a discrete set of values of the control variable, the discrete set in each group of the third set of groups extending over only a portion of the range from V_{lowest} to $V_{highest}$, the third analysis module providing simultaneous exchange of information between all values of the

control variable within a group of the third set of groups, each group in the third set of groups being distinct from each group in the first and second sets of groups.

8. A computer system according to claim 7, wherein the first, second and third analysis modules are formed from a common module of computer software.

9. A computer system according to claim 4, wherein the control variable comprises one of temperature, density, volume and pressure.

10. A computer system according to claim 4, wherein the physical system is a nanoscale structure.

11. A computer system for simulating a physical system using a control variable, comprising:

a first analysis module for computing data representative of the physical system over a range of temperatures, the range extending from a lowest value T_{lowest}, to a highest value, $T_{highest}$, the first analysis module individually computing data representative of the physical system within each of a first set of groups, each group in the first set including a discrete set of temperature values, the discrete set in each group of the first set of groups extending over only a portion of the range from T_{lowest} to $T_{highest}$, the first analysis module providing simultaneous exchange of information between all temperature values within a group of the first set of groups;

a second analysis module for computing data representative of the physical system over the temperature range extending from T_{lowest} to $T_{highest}$, the second analysis module individually computing data representative of the physical system within each of a second set of groups, each group in the second set including a discrete set of temperature values, the discrete set in each group of the second set of groups extending over only a portion of the range from T_{lowest} to $T_{highest}$, the second analysis module providing simultaneous exchange of information between all temperature values within a group of the second set of groups, each group in the second set of groups being distinct from each group in the first set of groups; and

a swapping module for communicating data such that the second analysis module can receive data computed by the first analysis module for a group in the first set of groups, the second analysis module using that data received from the first analysis module for computing data representative of the physical system for a group in the second set of groups.

12. A computer system according to claim 11, wherein the physical system is a nanoscale structure.

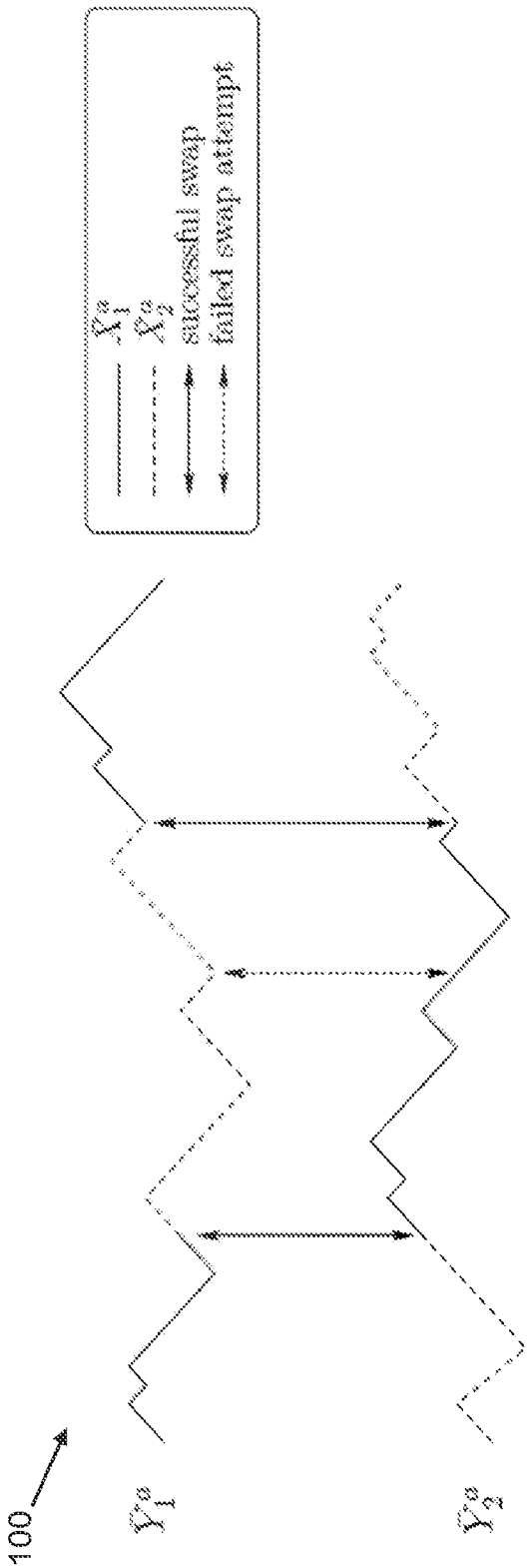

FIG. 1

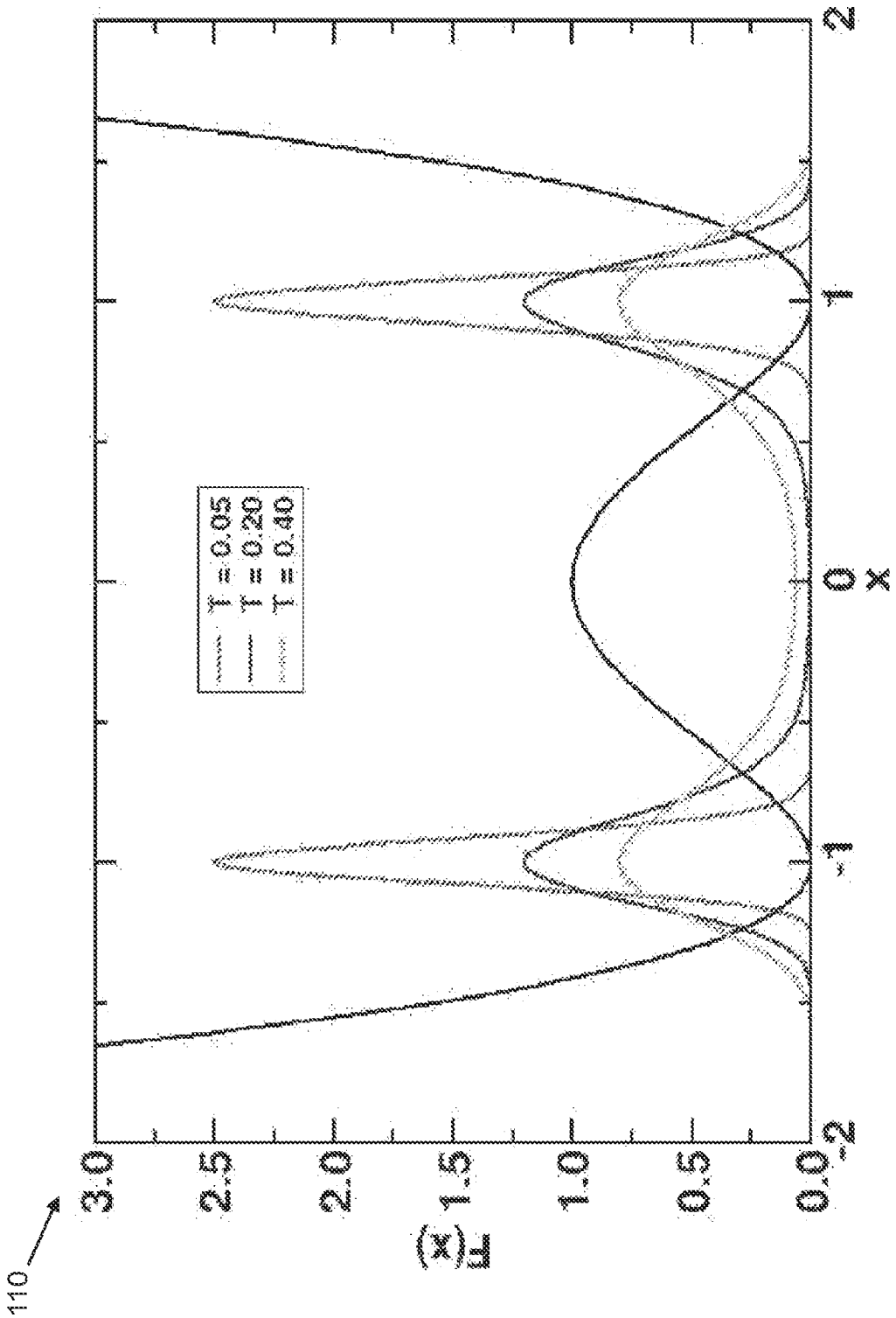

FIG. 1A

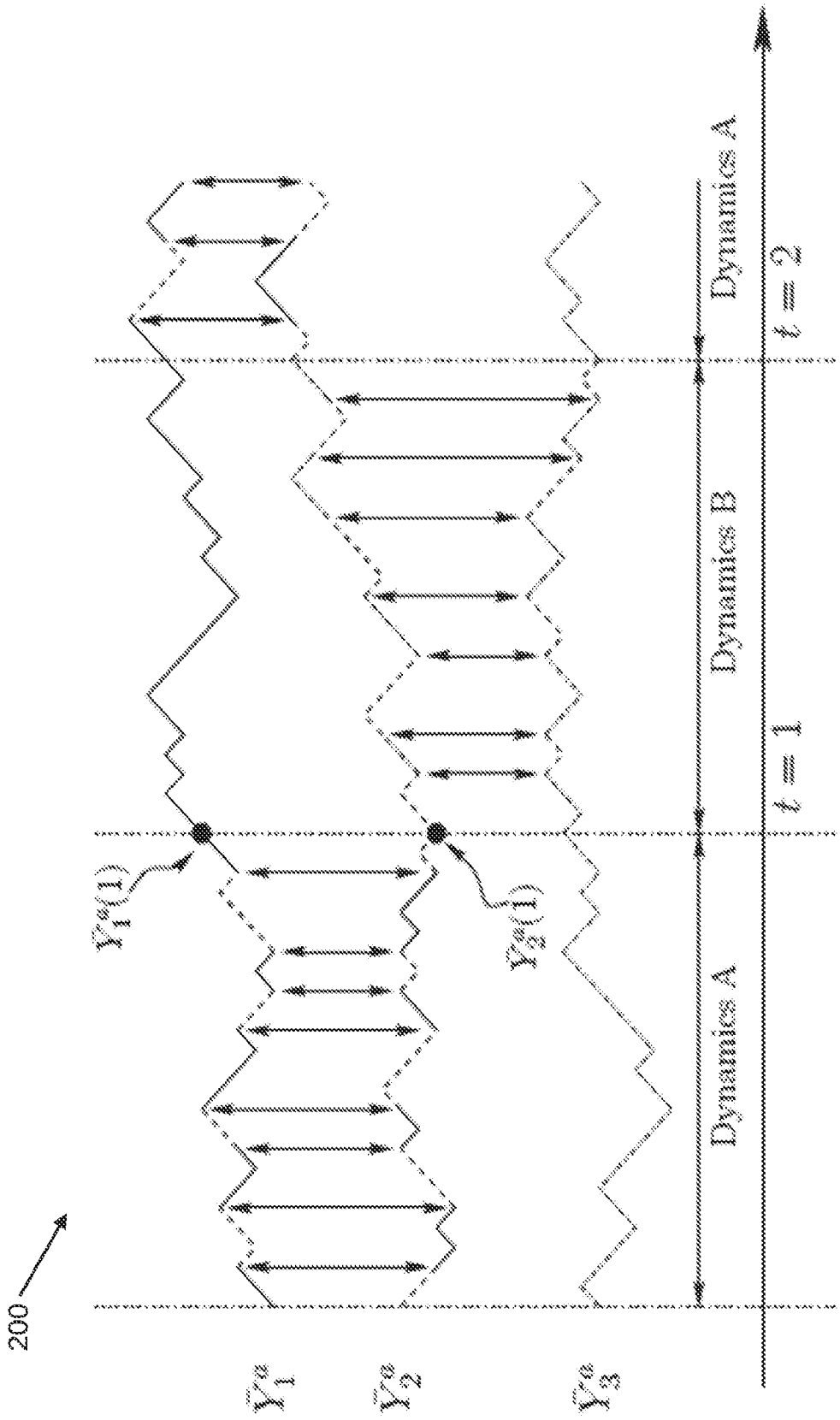
200
Y_1^a
Y_2^a
Y_3^a
$Y_1^a(1)$
$Y_2^a(1)$
Dynamics A
Dynamics B
Dynamics A
$t = 1$
$t = 2$

FIG. 2

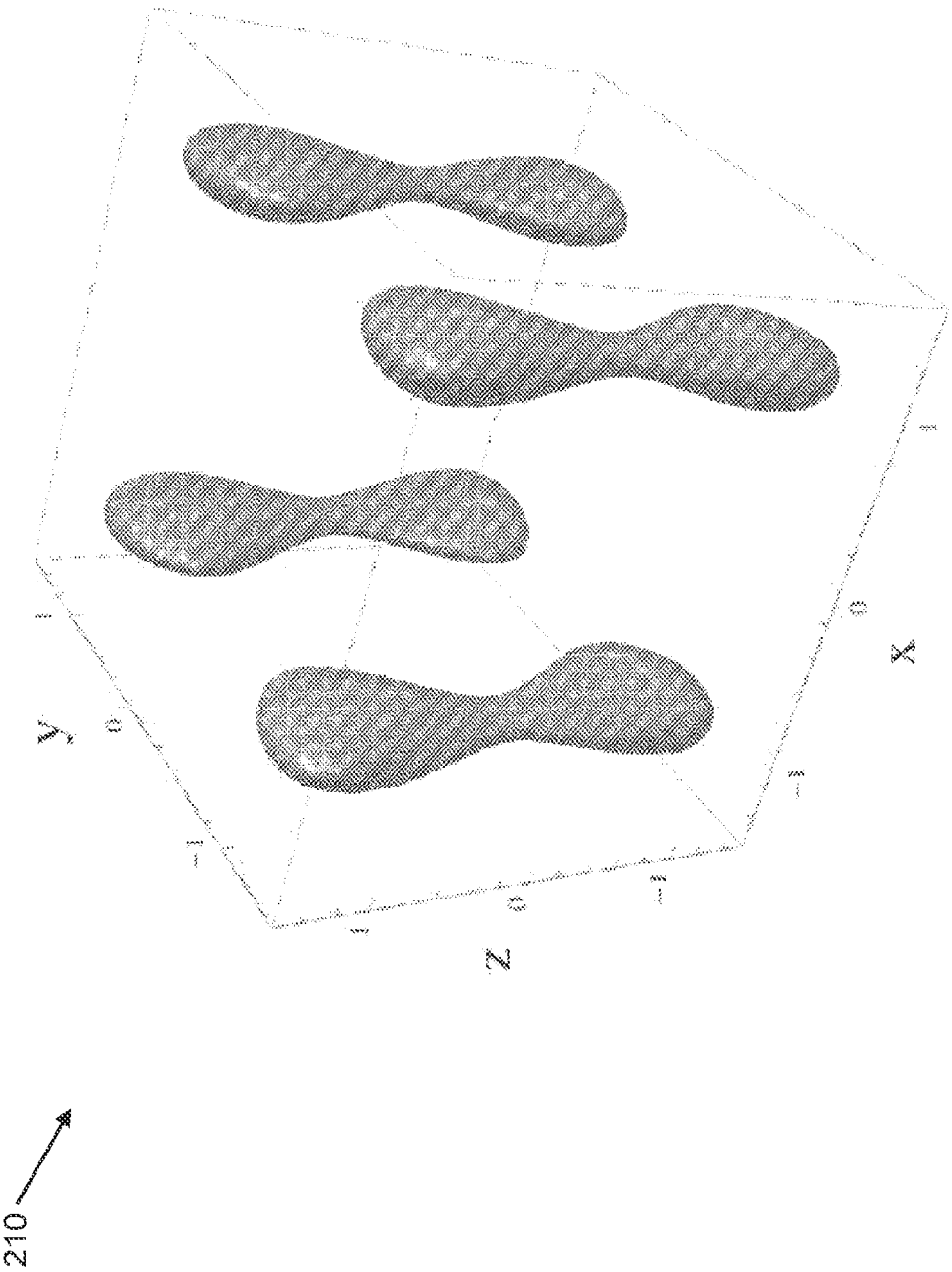

FIG. 2A

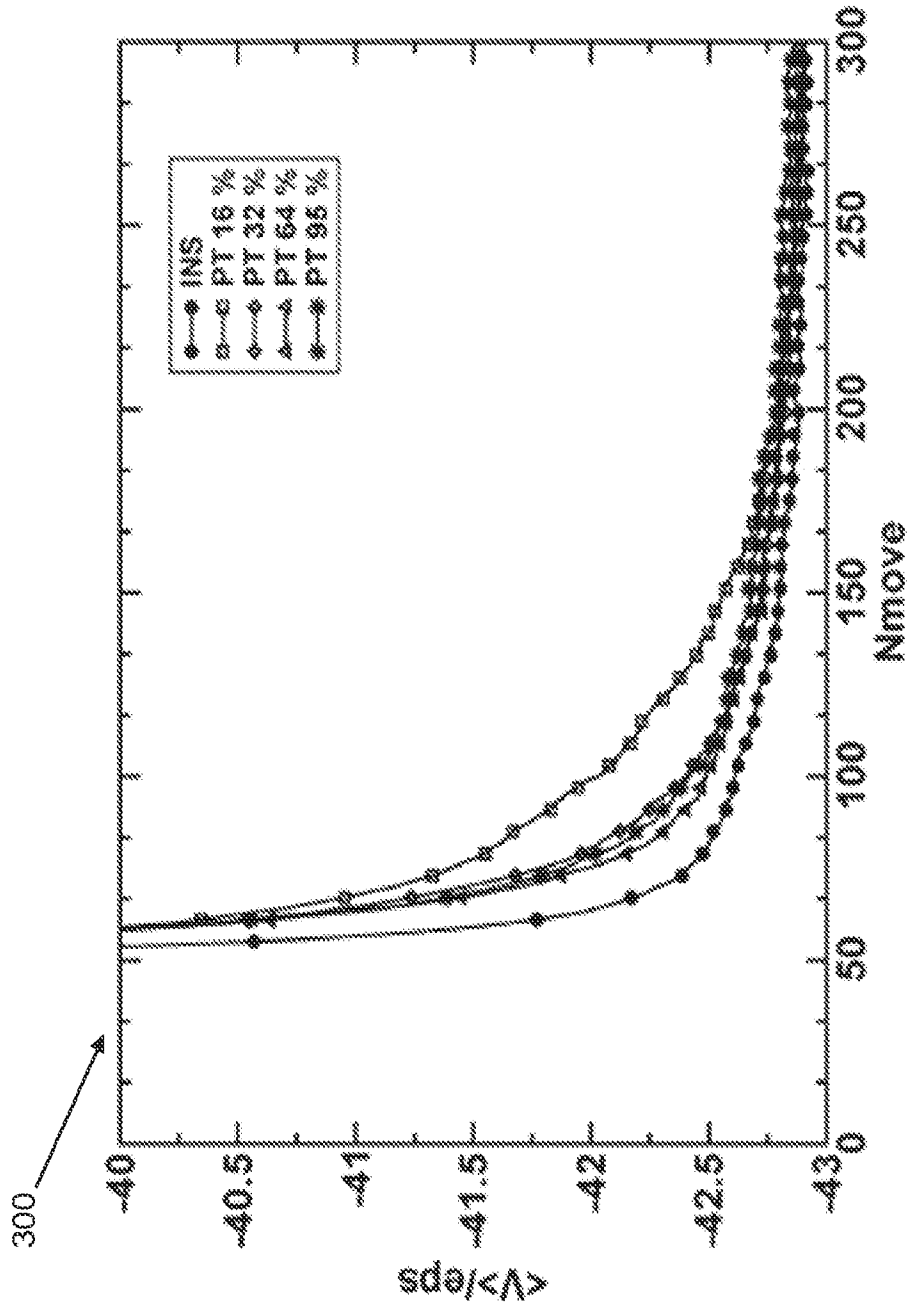
INS
PT 16 %
PT 32 %
PT 64 %
PT 95 %
<V>/eps
Nmove
300
-40
-40.5
-41
-41.5
-42
-42.5
-43
0
50
100
150
200
250
300

FIG. 3

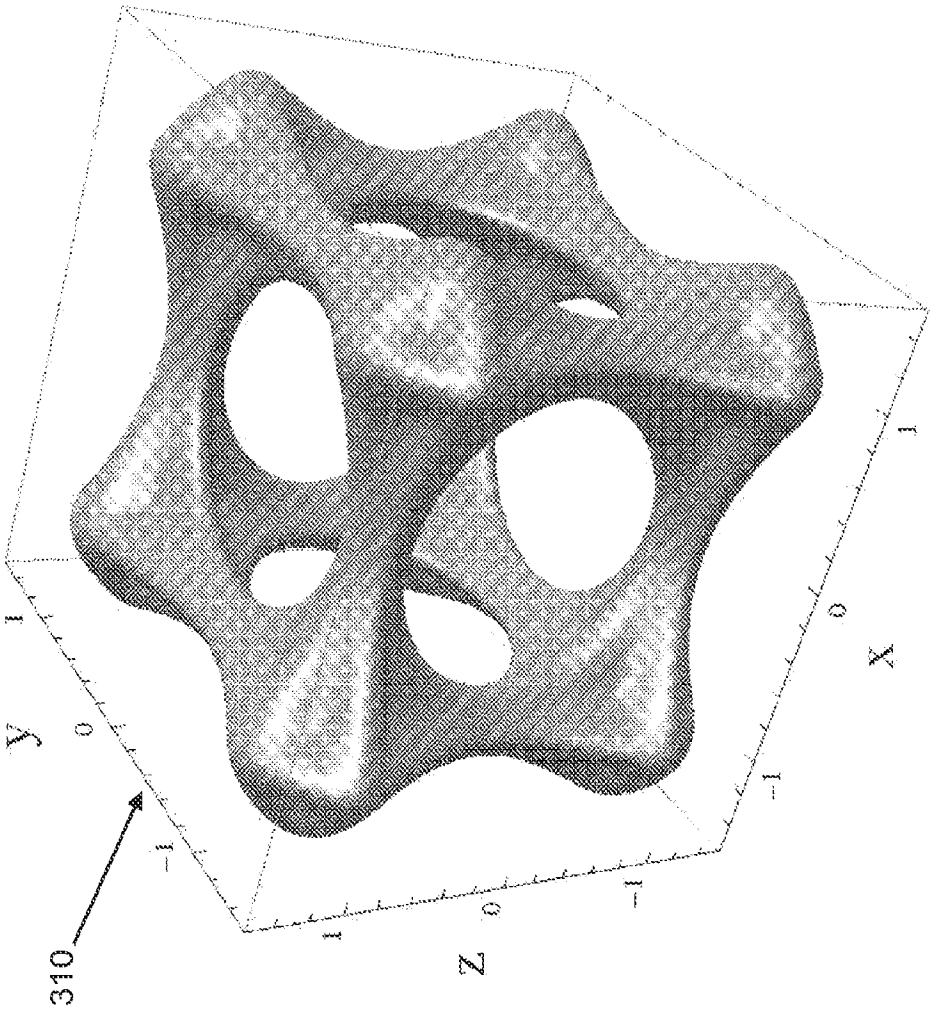

FIG. 3A

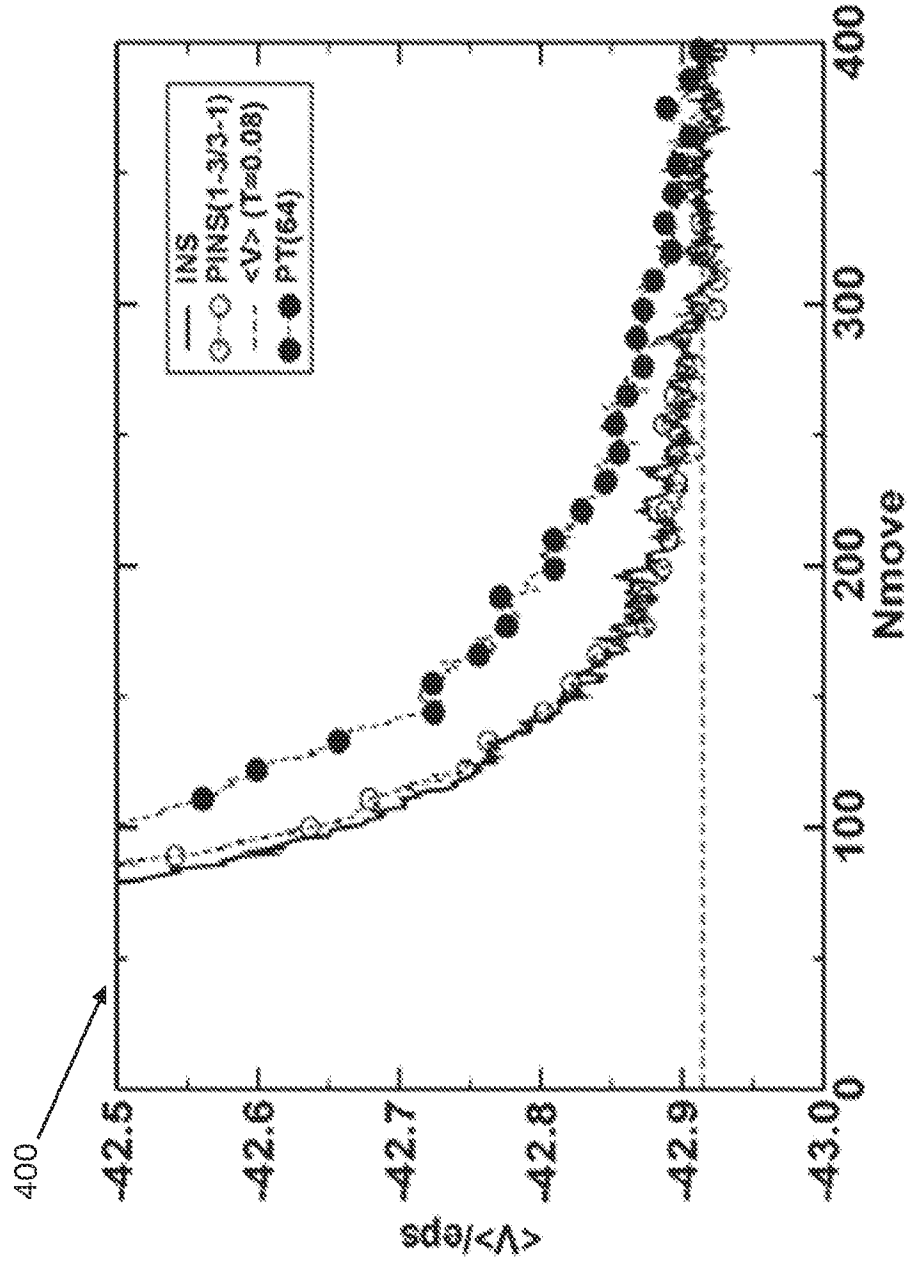
INS
PINS(1-3/3-1)
<V> (T=0.08)
PT(64)
400
-42.5
-42.6
-42.7
-42.8
-42.9
-43.0
<V>/eps
0
100
200
300
400
Nmove

FIG. 4

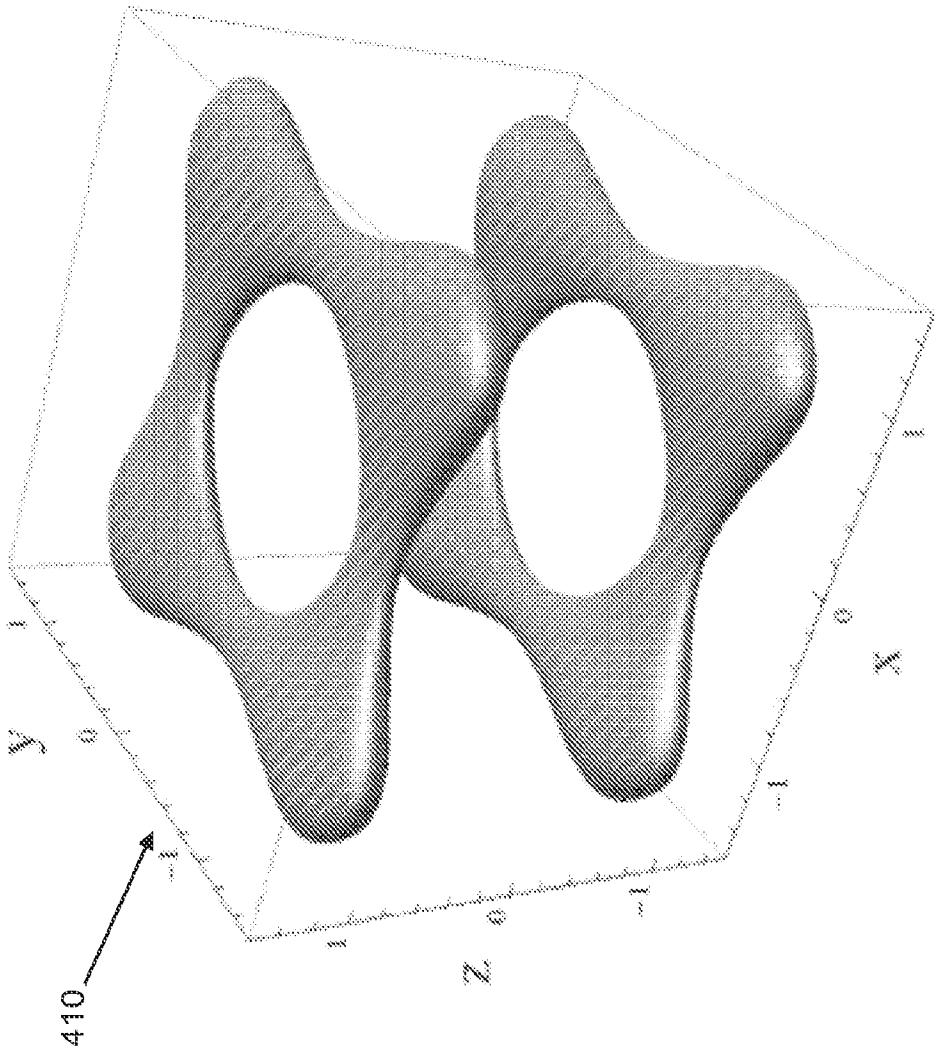

FIG. 4A

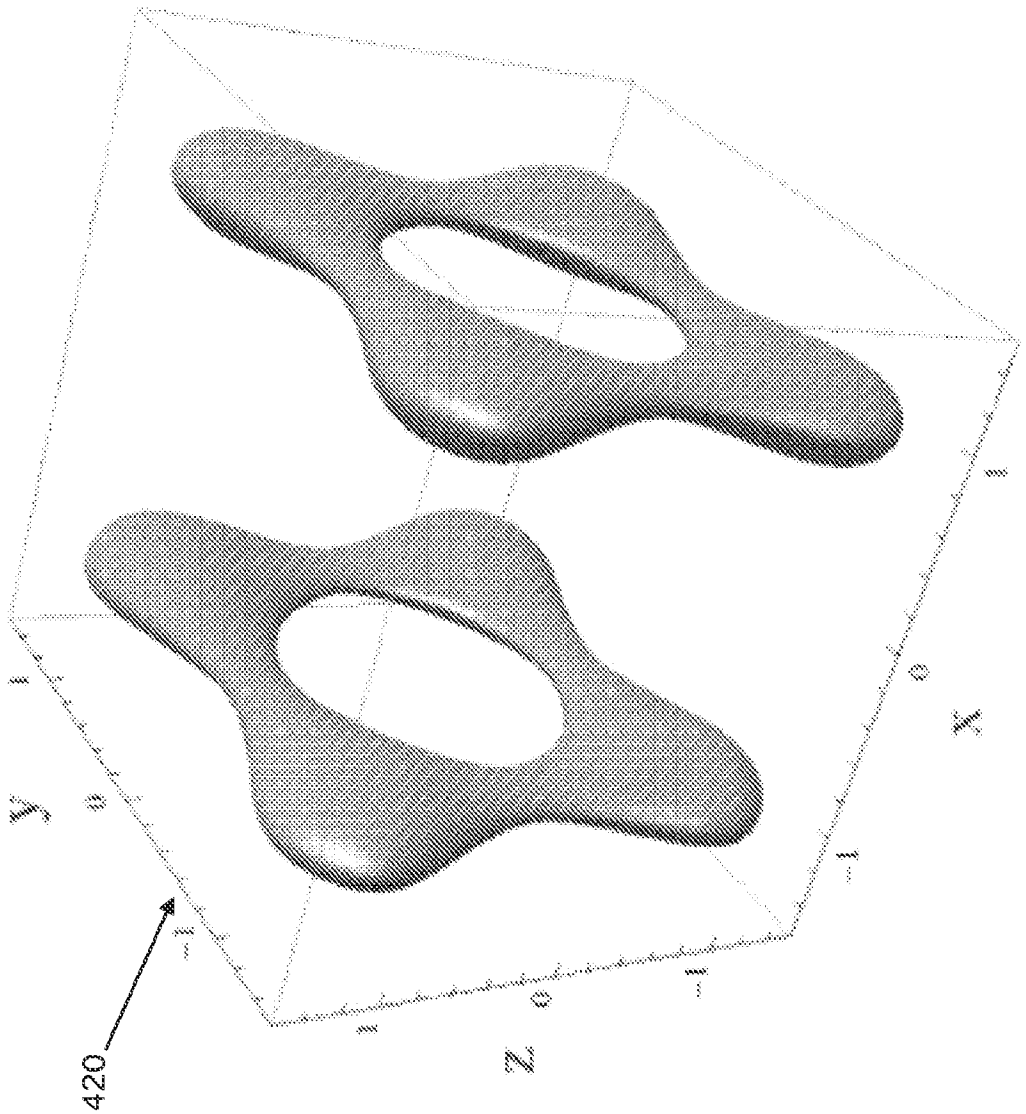

FIG. 4B

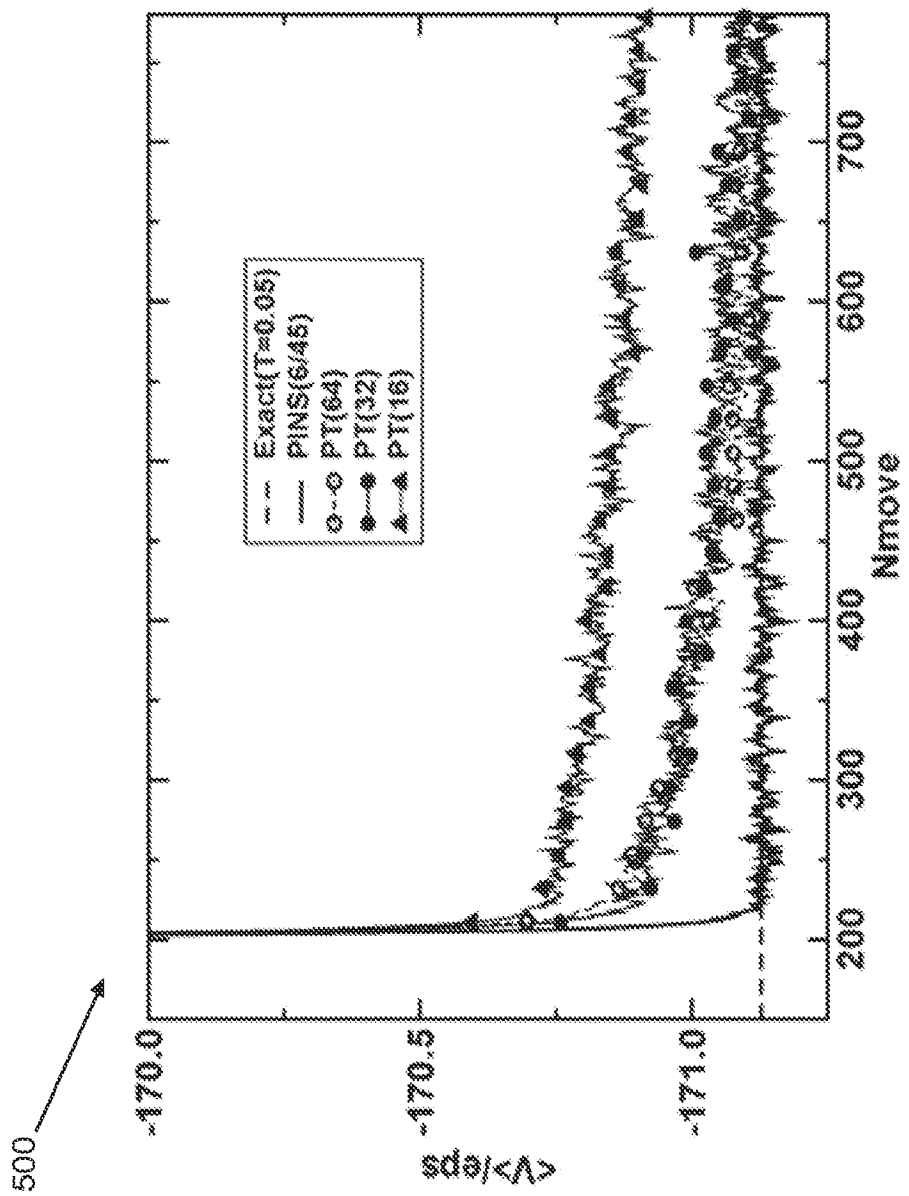
500
-170.0
-170.5
-171.0
<V>/eps
200
300
400
500
600
700
Nmove
Exact(T=0.05)
PINS(6/45)
PT(64)
PT(32)
PT(16)

FIG. 5

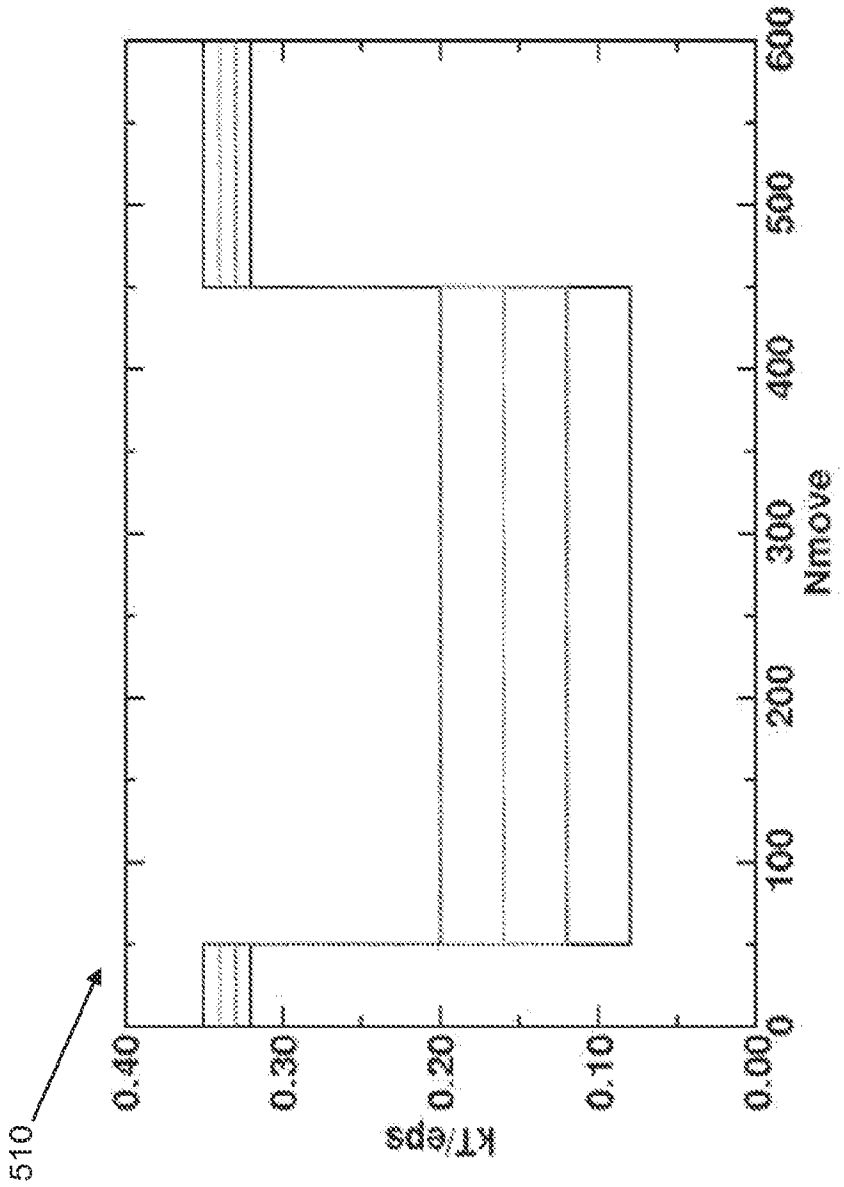

FIG. 5A

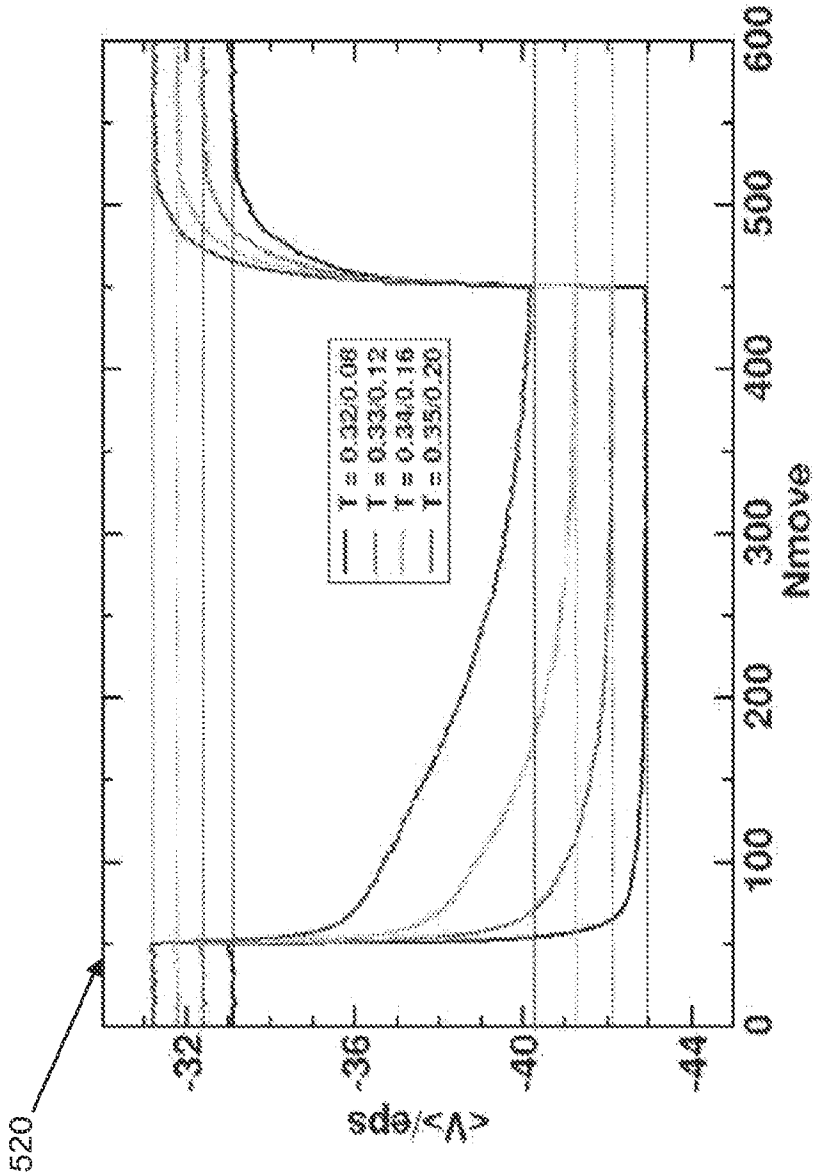
520
-32
-36
-40
-44
<V>/eps
0
100
200
300
400
500
600
Nmove
T = 0.32/0.08
T = 0.33/0.12
T = 0.34/0.16
T = 0.35/0.20

FIG. 5B

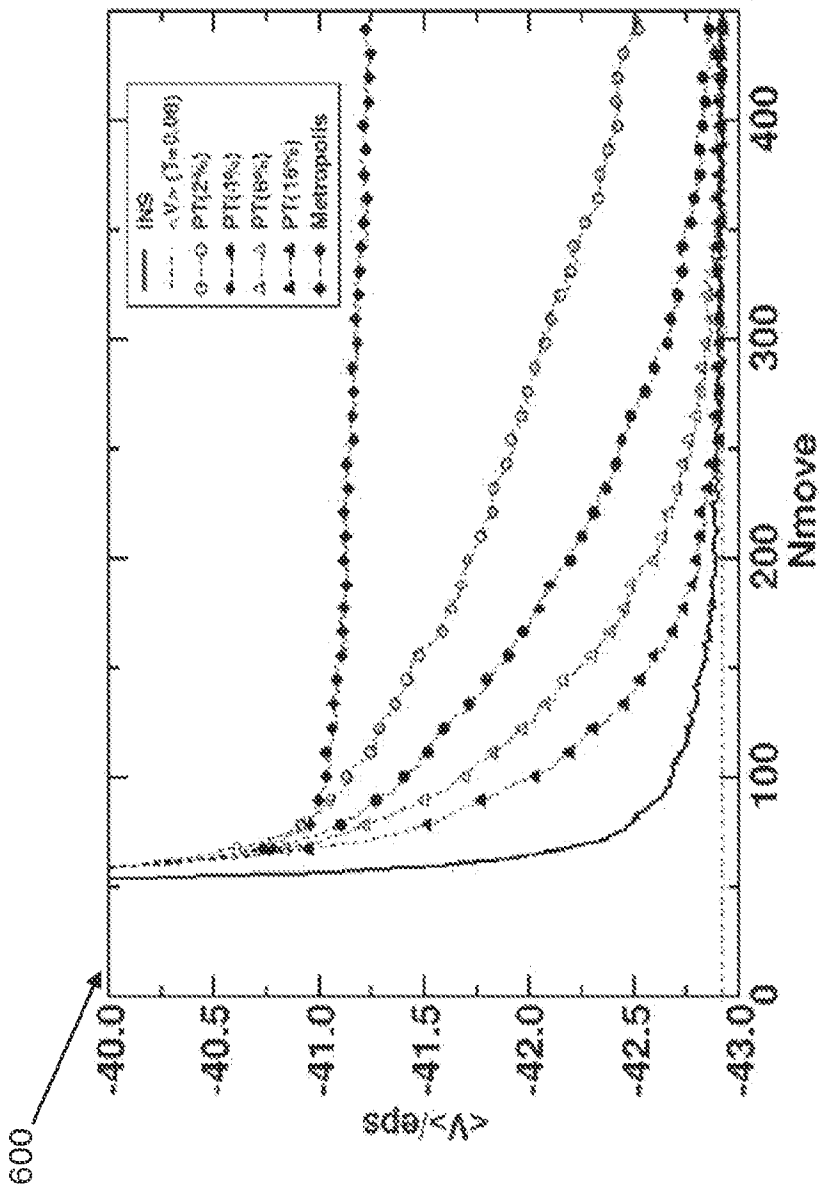
INS
Metropolis
-40.0
-40.5
-41.0
-41.5
-42.0
-42.5
-43.0
<V>/eps
0
100
200
300
400
Nmove
600

FIG. 6

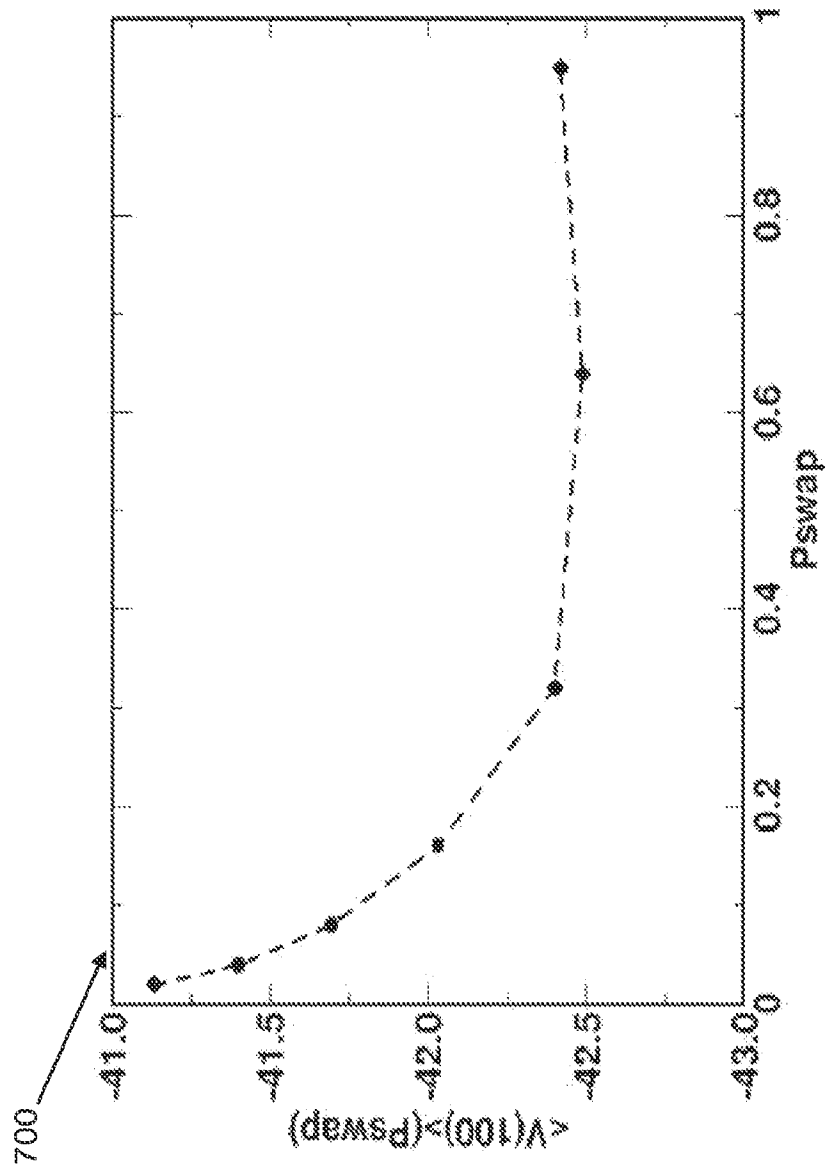

FIG. 7

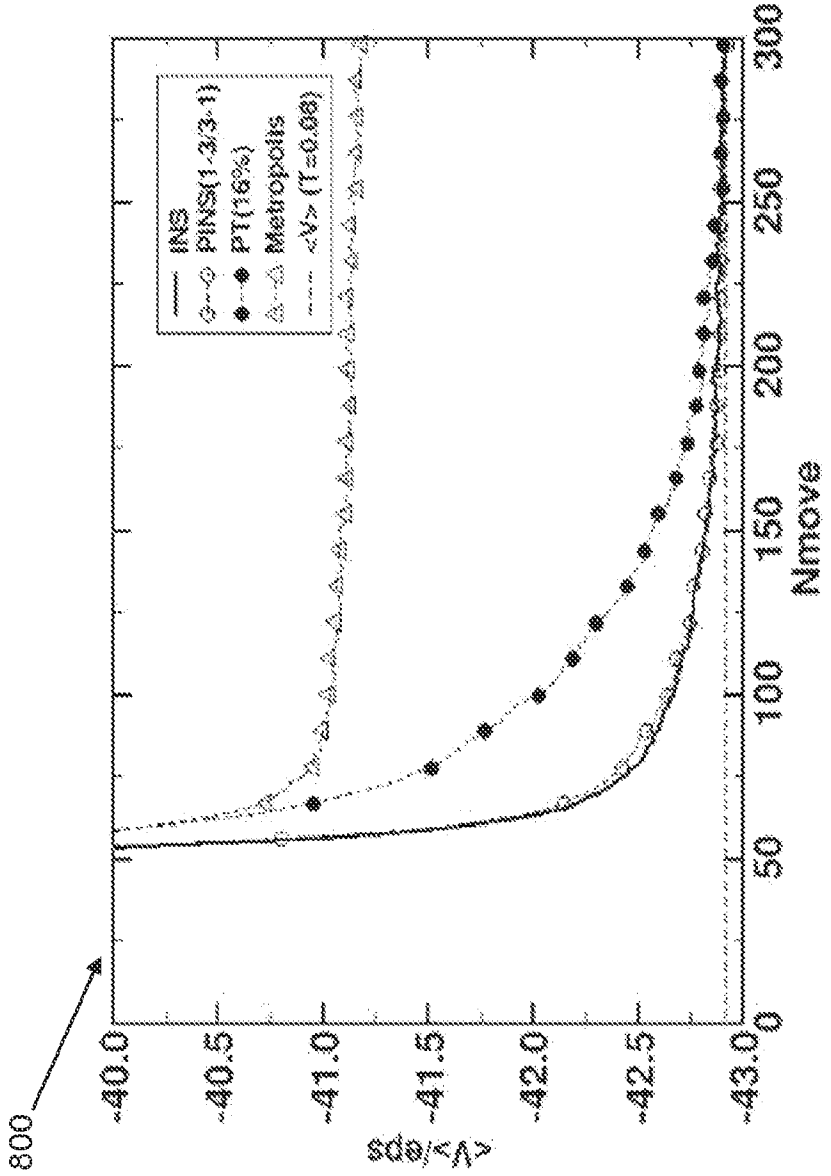
800
-40.0
-40.5
-41.0
-41.5
-42.0
-42.5
-43.0
<V>/eps
0
50
100
150
200
250
300
Nmove
INS
PINS(1-3/3-1)
PT(16%)
Metropolis
<V> (T=0.68)

FIG. 8

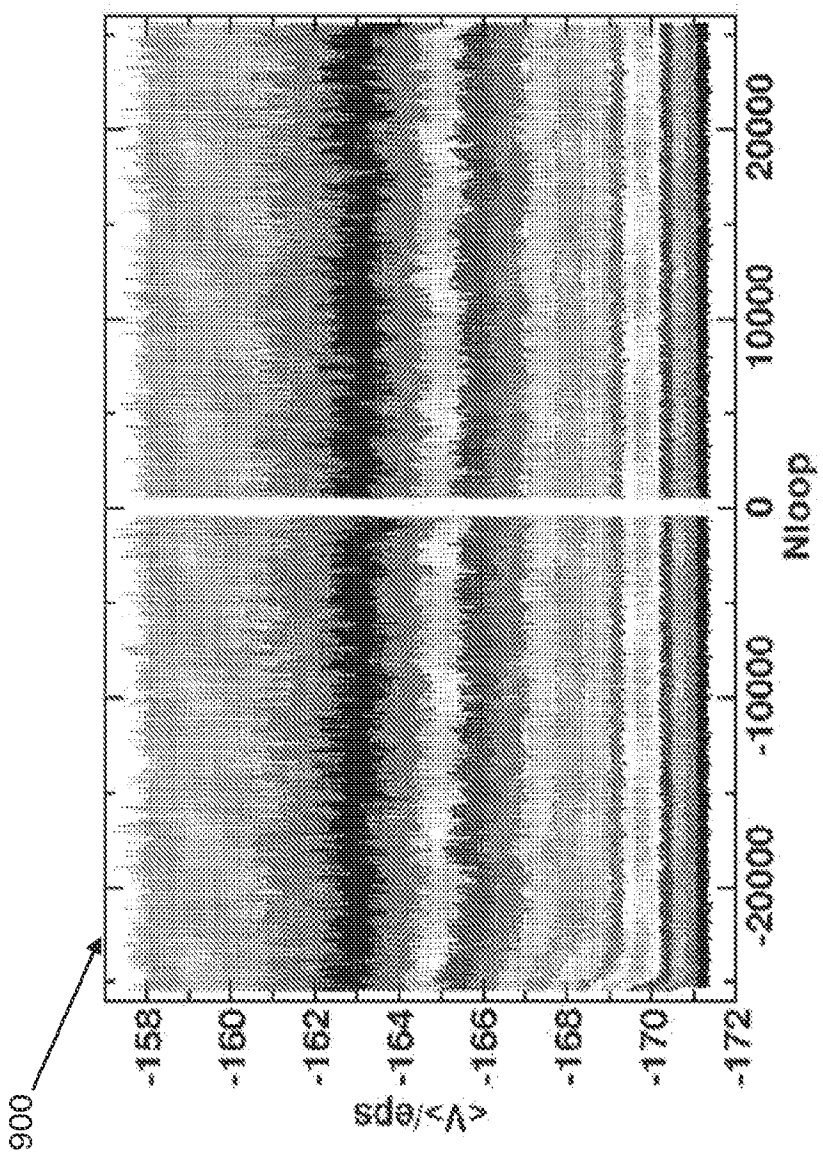
900
-158
-160
-162
-164
-166
-168
-170
-172
<V>/eps
-20000
-10000
0
10000
20000
Nloop

FIG. 9

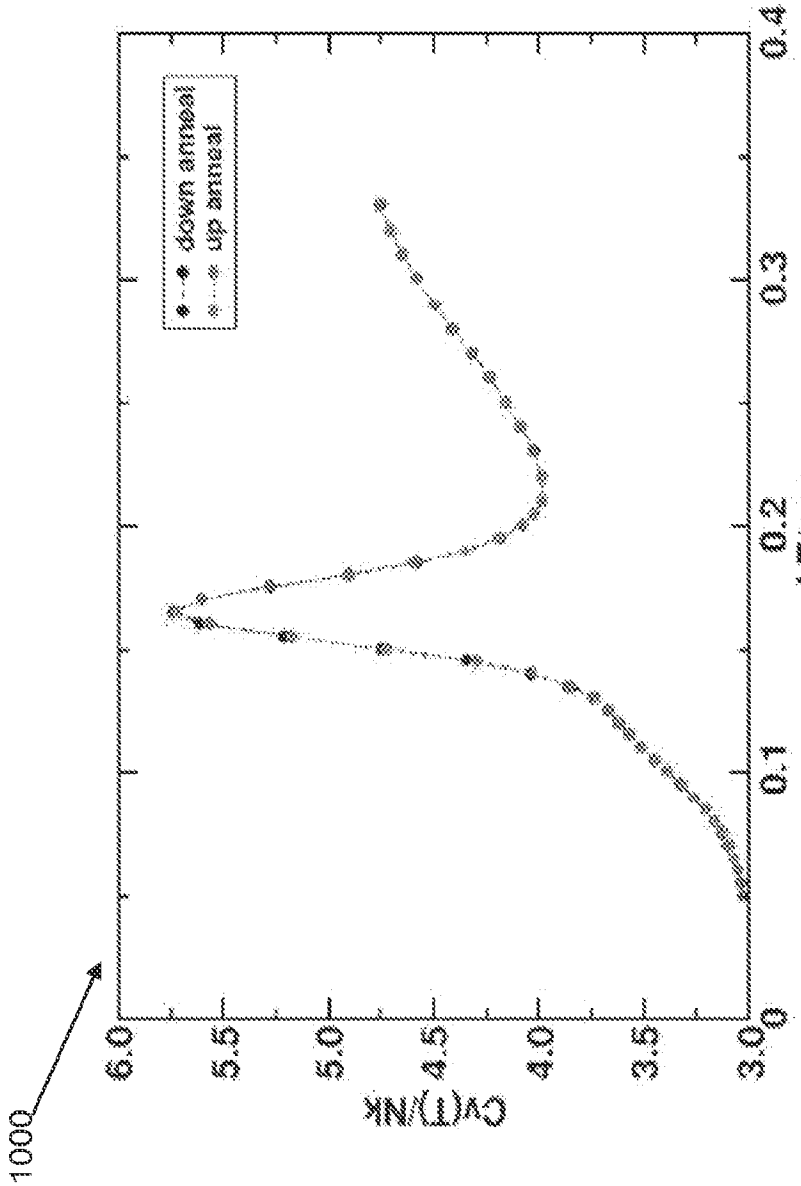
1000
down anneal
up anneal
Cv(T)/Nk
kT/eps
6.0
5.5
5.0
4.5
4.0
3.5
3.0
0
0.1
0.2
0.3
0.4

FIG. 10

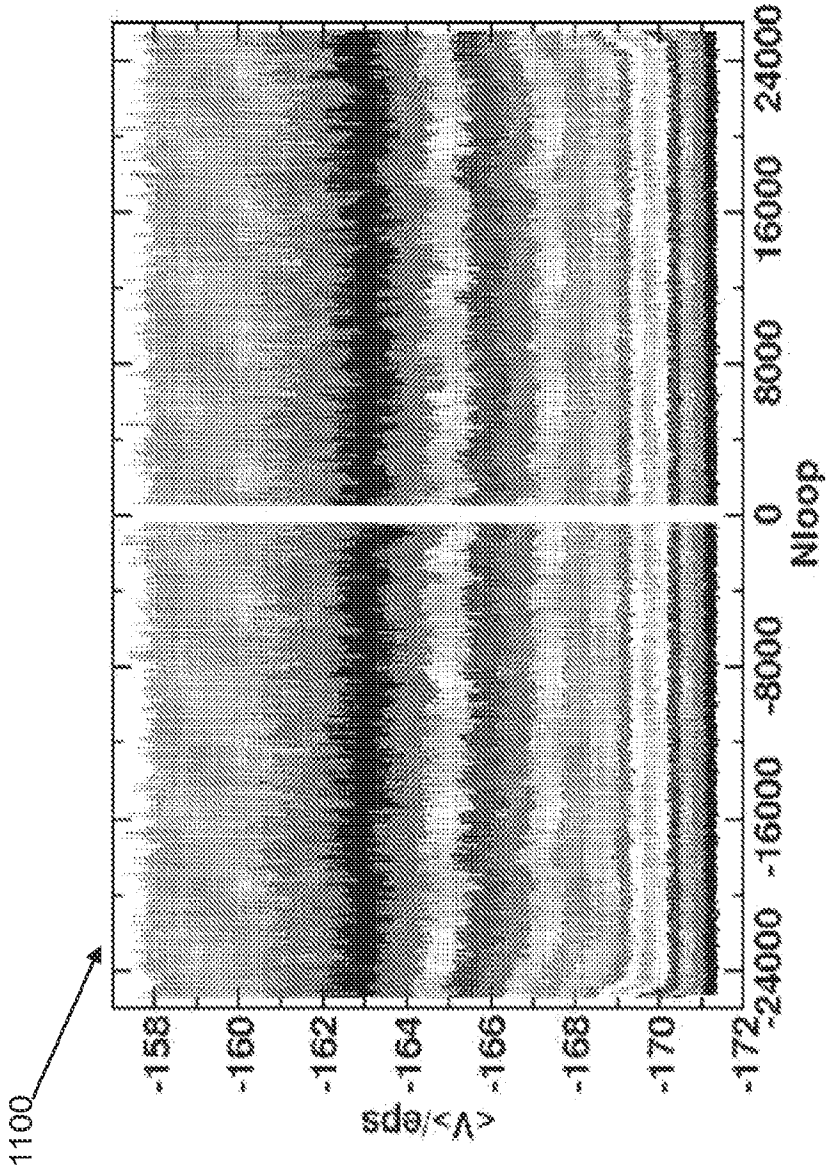
1100
-158
-160
-162
-164
-166
-168
-170
-172
<V>/eps
-24000
-16000
-8000
0
8000
16000
24000
Nloop

FIG. 11

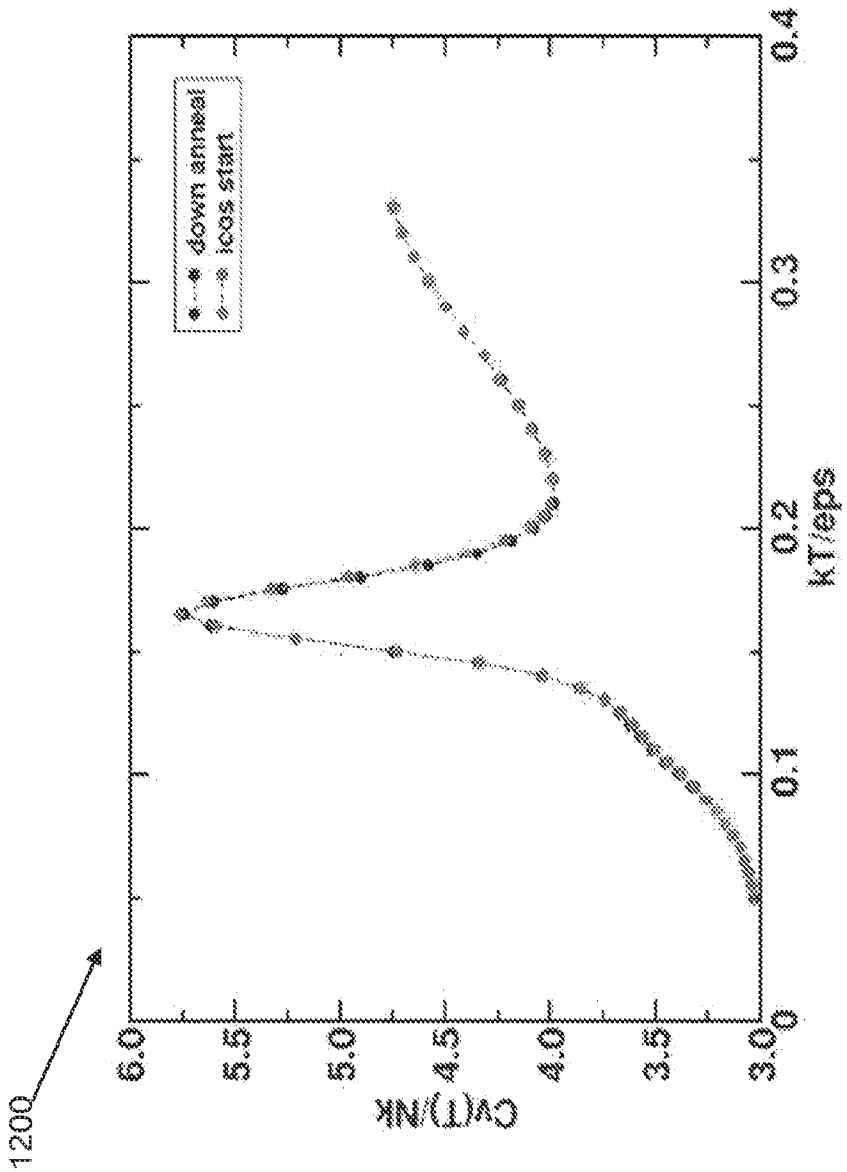
1200
6.0
5.5
5.0
4.5
4.0
3.5
3.0
Cv(T)/Nk
0.1
0.2
0.3
0.4
kT/eps
down anneal
ico start

FIG. 12

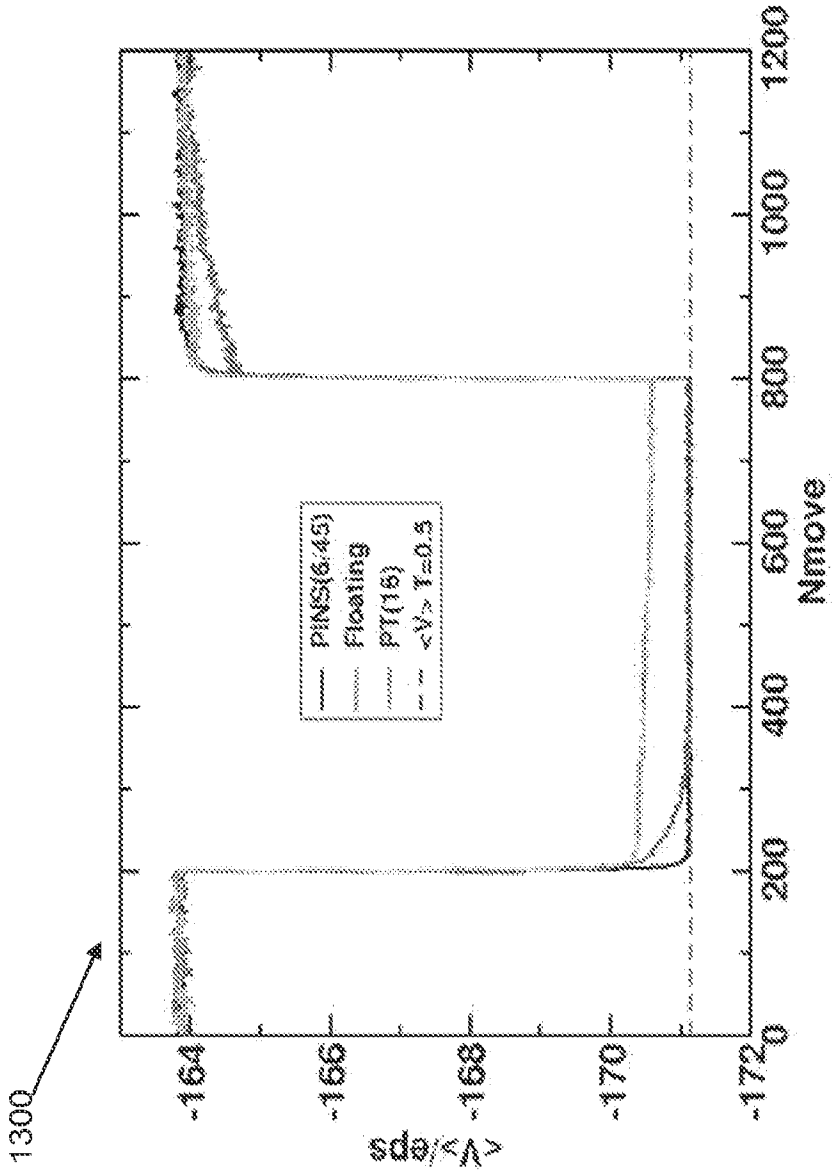
1300
<V>/eps
-164
-166
-168
-170
-172
0
200
400
600
800
1000
1200
Nmove
PINS(6.45)
Floating
PT(16)
<V> T=0.5

FIG. 13

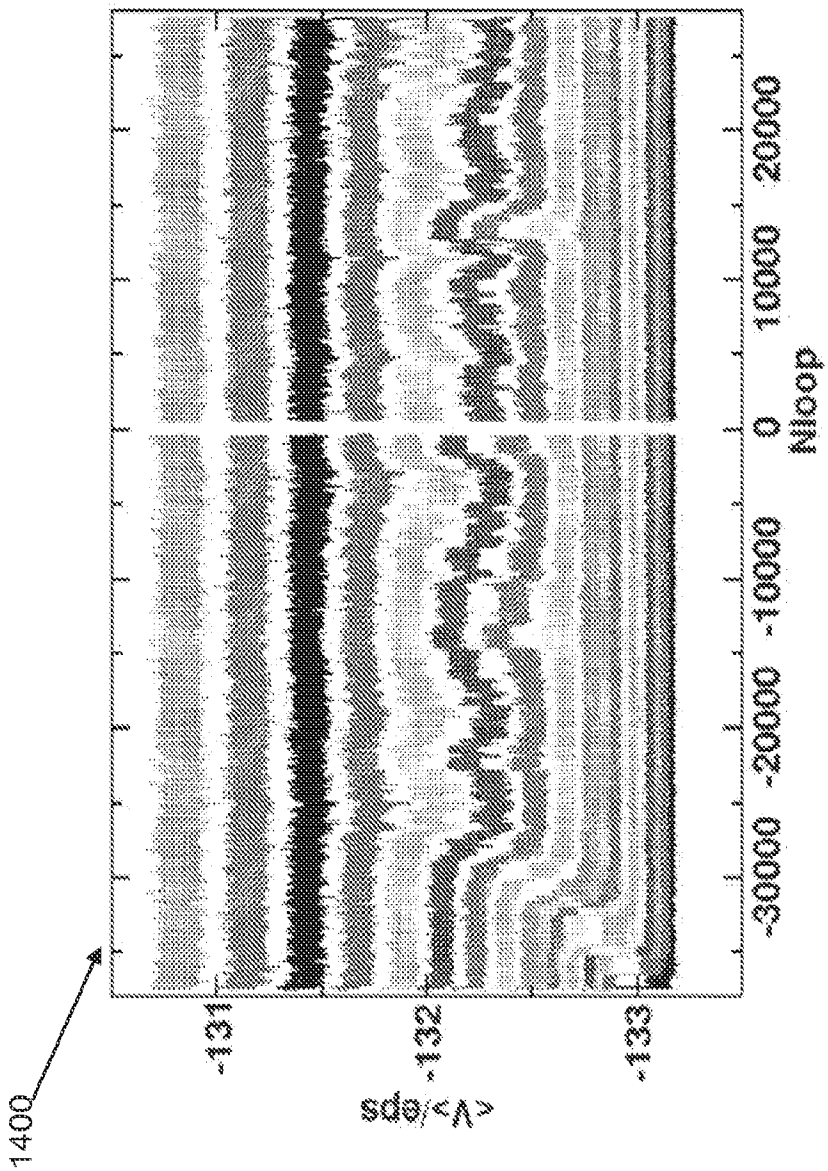

FIG. 14

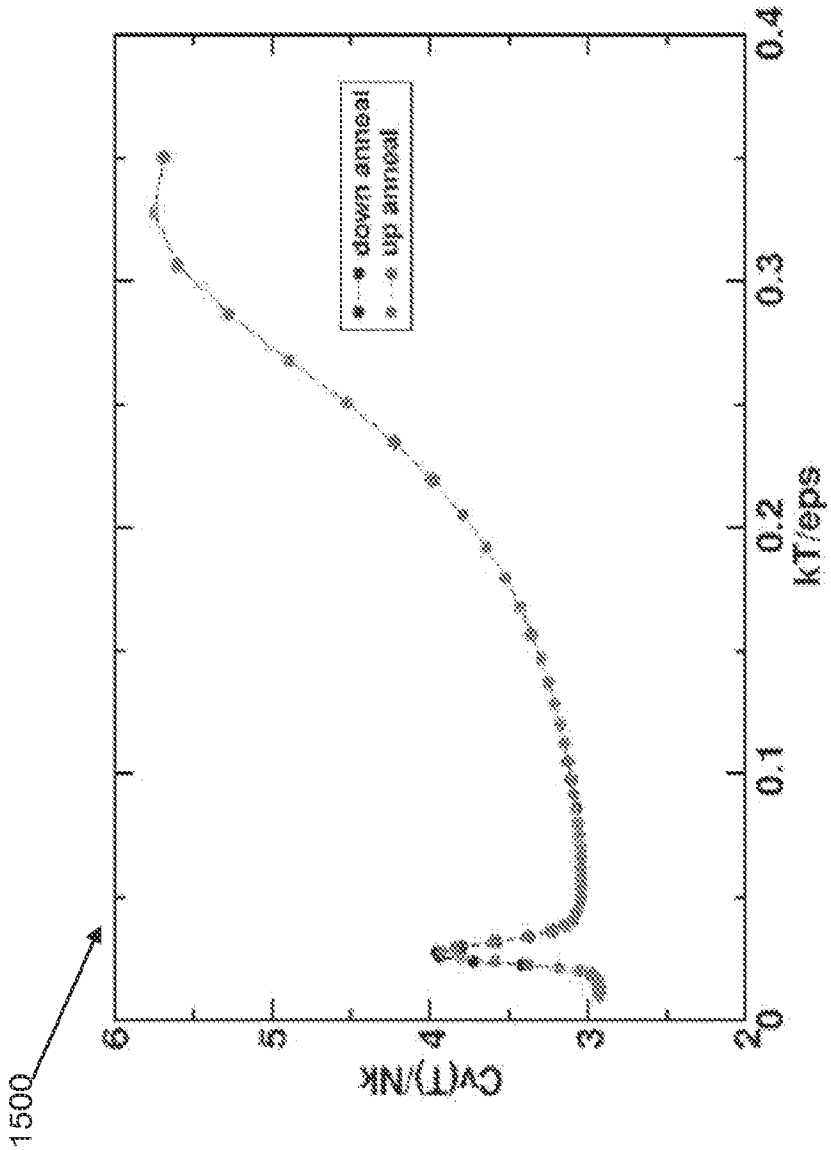

FIG. 15

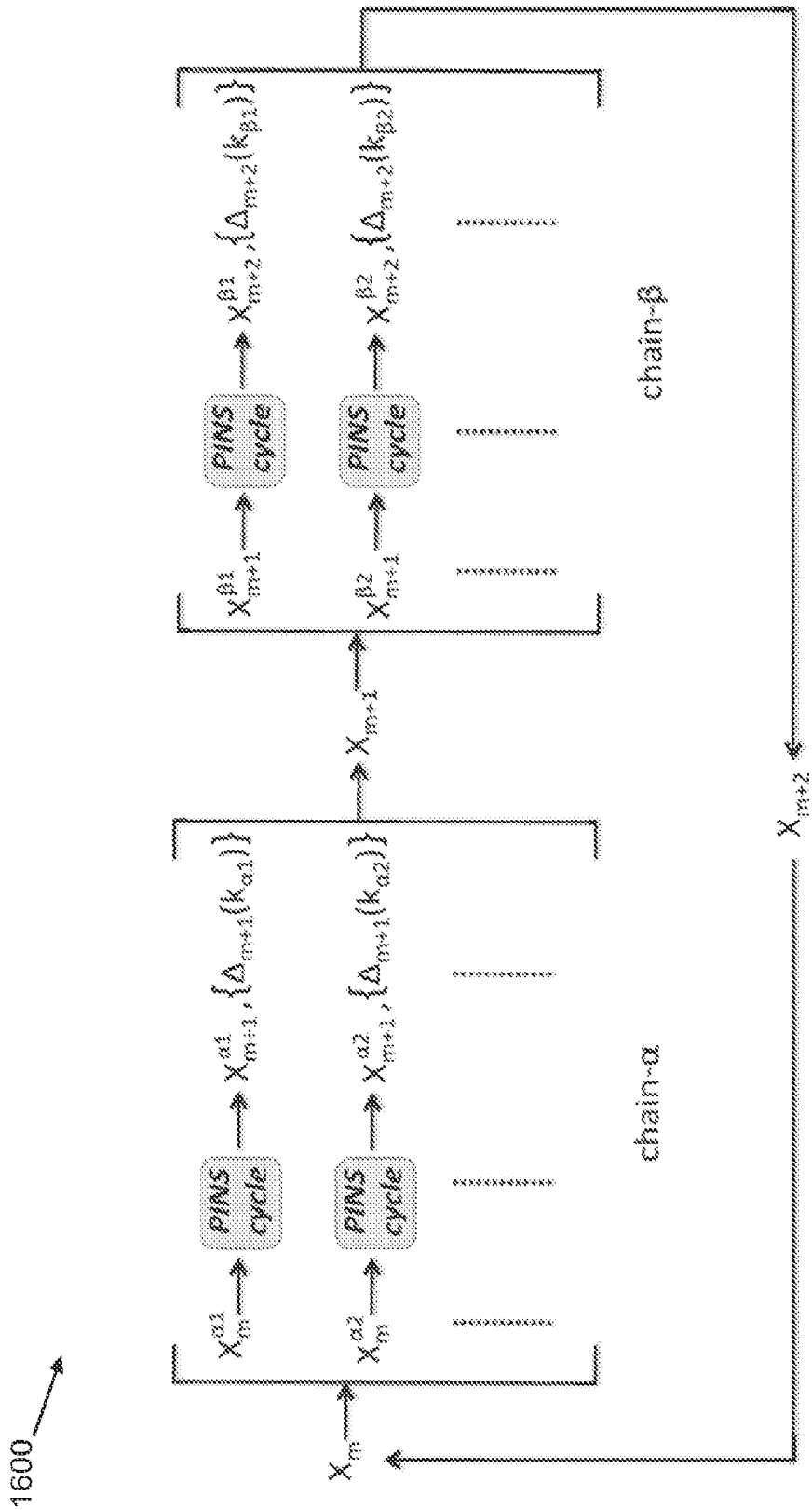
1600
X_m
$X_m^{\alpha 1}$
$X_m^{\alpha 2}$
PINS cycle
PINS cycle
$X_{m+1}^{\alpha 1}, \{\Delta_{m+1}(k_{\alpha 1})\}$
$X_{m+1}^{\alpha 2}, \{\Delta_{m+1}(k_{\alpha 2})\}$
chain-α
X_{m+1}
$X_{m+1}^{\beta 1}$
$X_{m+1}^{\beta 2}$
PINS cycle
PINS cycle
$X_{m+2}^{\beta 1}, \{\Delta_{m+2}(k_{\beta 1})\}$
$X_{m+2}^{\beta 2}, \{\Delta_{m+2}(k_{\beta 2})\}$
chain-β
X_{m+2}

FIG. 16

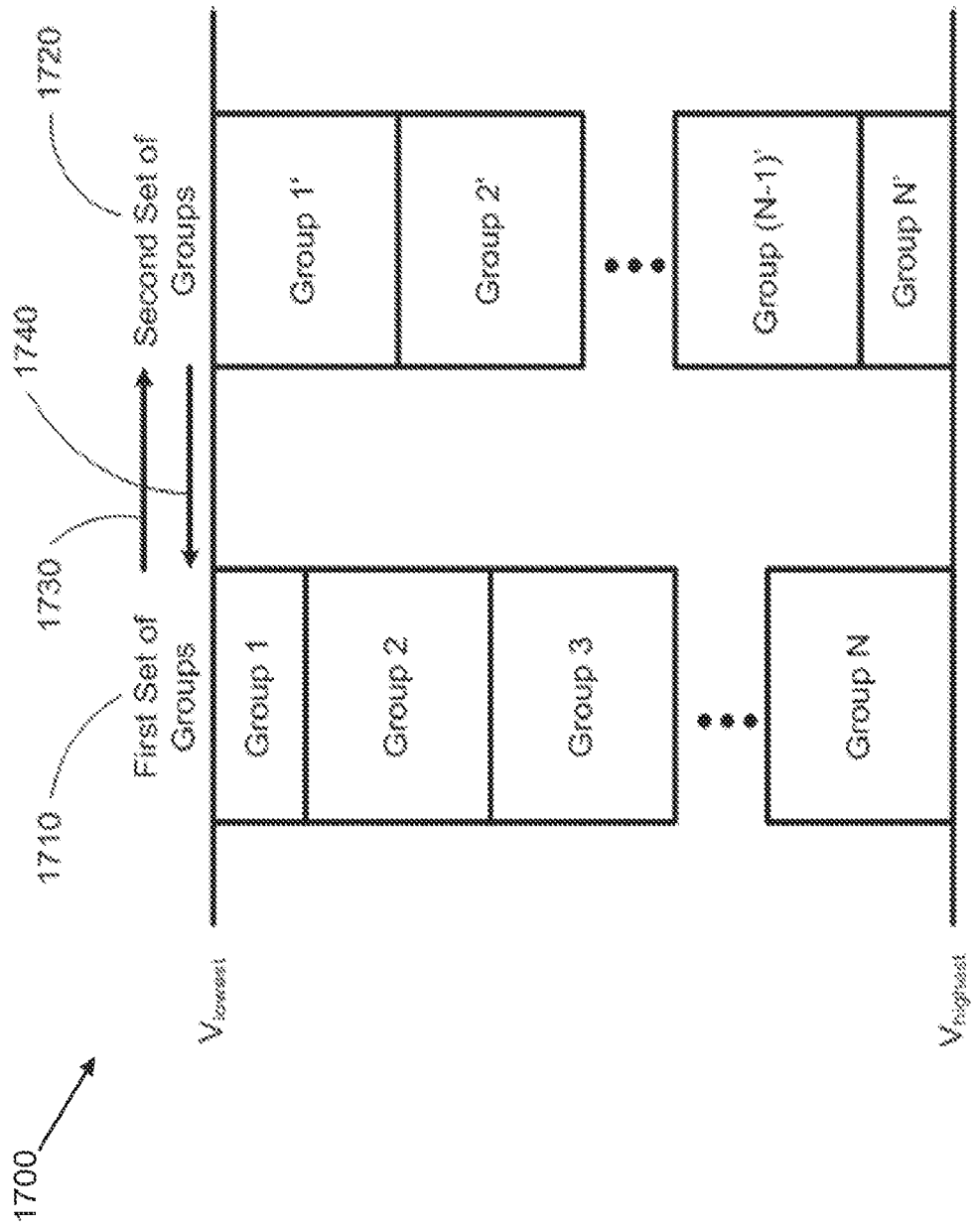
1700
1710
First Set of Groups
1720
Second Set of Groups
1730
1740
Group 1
Group 2
Group 3
Group N
Group 1'
Group 2'
Group (N-1)'
Group N'
V_{lowest}
$V_{highest}$

FIG. 17

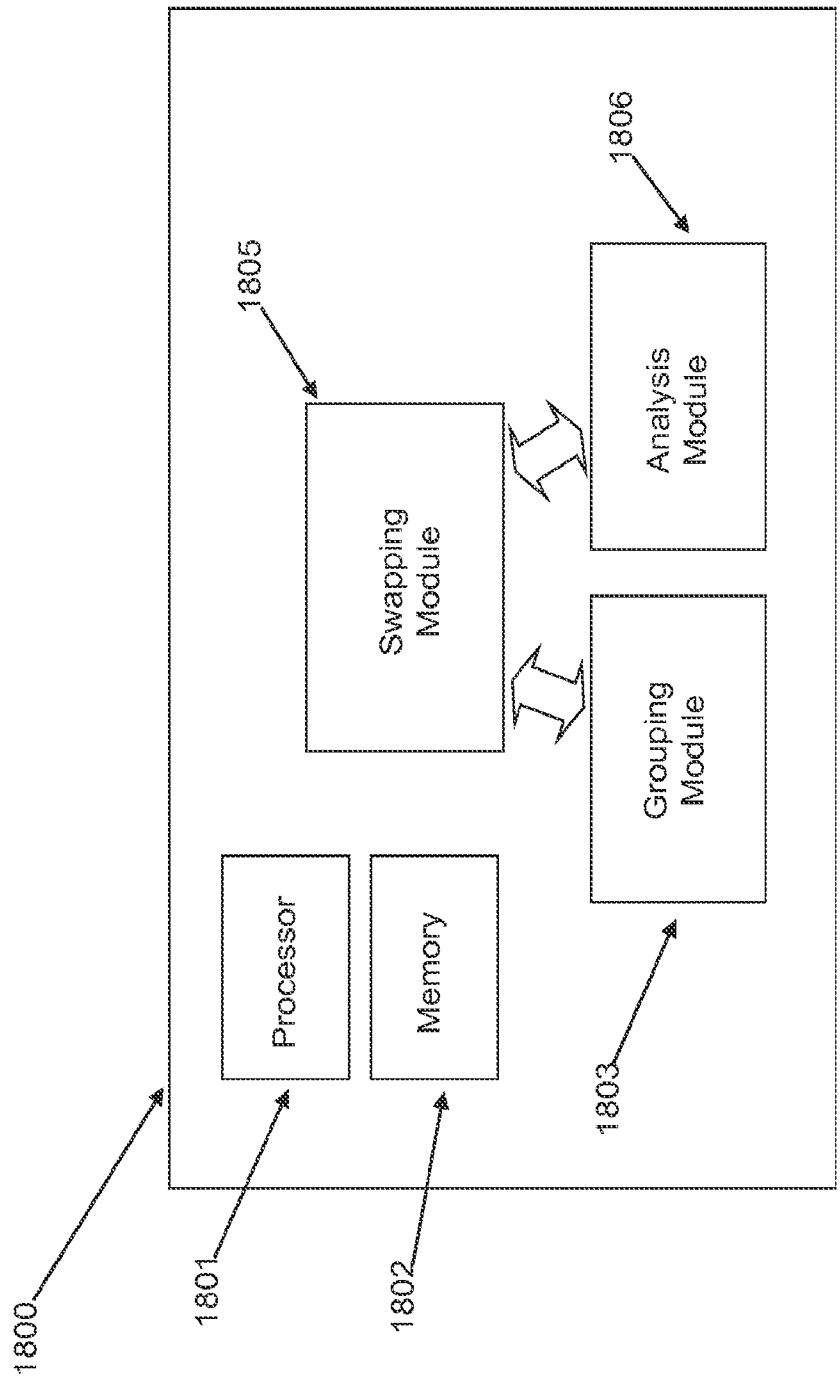
1800
1801
1802
1803
1805
1806
Processor
Memory
Swapping Module
Grouping Module
Analysis Module

FIG. 18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/45779

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 17/50 (2012.01)
USPC - 703/16
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): G06F 17/50 (2012.01)
USPC: 703/16

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 703/2, 17; 706/46 (view search terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
DialogWeb (347,348,349,351,371,652,654); Google Scholar; Google Web; Google Patents.
Search Terms: SIMULATE, GROUP, DIVIDE, LOWEST, HIGHEST, DENSITY, PRESSURE, VOLUME, TEMPERATURE, ANALYZE, SWAP, EXCHANGE, SIMULTANEOUS, SAME TIME, CONCURRENT, COMPUTE, NANO, NANOSCALE, SOFTWARE, PROGRAM

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	VAN ERP. "Dynamical Rare event simulation techniques for equilibrium and non-equilibrium systems." In Advances in Chemical Physics [online]. Published December, 2010. [retrieved on 2012-09-05] Retrieved from the Internet <URL: http://arxiv.org/pdf/1101.0927.pdf > entire document, especially Abstract; pg 1, para 1 to pg 3, para 2; pg 3, para 5 to pg 5, para 1; pg 11, para 3 to pg 12, para 1; pg 14, para 3; pg 20, para 1	1-12
L	CITATION depicting publication name: VAN ERP. "Dynamical Rare event simulation techniques for equilibrium and non-equilibrium systems." In Advances in Chemical Physics, pages 4189-4196 [online]. Published December, 2010. [retrieved on 2012-09-05] Retrieved from the Internet <URL: http://arxiv.org/pdf/1101.0927.pdf > entire document	1-12
Y	PREDESCU et al. "On the Efficiency of Exchange in Parallel Tempering Monte Carlo Simulations." In the Journal of Phsysical Chemistry B, 2005, vol 109, number 9, pages 4189-4196 [online]. Published October 28, 2004. [retrieved on 2012-09-05] Retrieved from the Internet. <URL: http://www.diggernet.net/fs_home/cciobanu/publications/ptmceff.pdf > entire document, especially Abstract; pg 4189, para 1; pg 4189, para 4 to pg 4190, para 1	1-12
Y	KUPPA et al. "Dynamics of poly(ethylene oxide) in nanoscale confinements: A computer simulations perspective." In Journal of Chemical Physics, vol. 118, number 7, pages 3421-3429 [online]. Published September 3, 2002. [retrieved on 2012-09-05] Retrieved from the Internet <URL: http://raman.plmsc.psu.edu/~manias/PDFs/PEO_jcp03.pdf > entire document, especially Abstract; pg 3421, para 1; pg 3423, para 2	1-12

☒ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
05 September 2012 (05.09.2012)	24 SEP 2012

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/45779

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 7,891,818 B2 (CHRISTENSEN et al.) 22 February 2011 (22.02.2011) entire document	1-12
A	US 2009/0204374 A1 (HILL et al.) 13 August 2009 (13.08.2009) entire document	1-12
A	US 2008/0124726 A1 (MONFORTE) 29 May 2008 (29.05.2008) entire document	1-12
A	US 5,825,660 A (CAGAN et al.) 20 October 1998 (20.10.1998) entire document	1-12