Infinite swapping in curved spaces

E. Curotto[1,a)] and Massimo Mella[2,b)]
[1]*Department of Chemistry and Physics, Arcadia University, Glenside, Pennsylvania 19038-3295, USA*
[2]*Dipartimento di Scienza ed Alta Tecnologia, via Valleggio 11, Università degli studi dell'Insubria, 22100 Como, Italy*

(Received 19 March 2013; accepted 9 December 2013; published online 3 January 2014)

We develop an extension of the infinite swapping and partial infinite swapping techniques [N. Plattner, J. D. Doll, P. Dupuis, H. Wang, Y. Liu, and J. E. Gubernatis, J. Chem. Phys. **135**, 134111 (2011)] to curved spaces. Furthermore, we test the performance of infinite swapping and partial infinite swapping in a series of flat spaces characterized by the same potential energy surface model. We develop a second order variational algorithm for general curved spaces without the extended Lagrangian formalism to include holonomic constraints. We test the new methods by carrying out NVT classical ensemble simulations on a set of multidimensional toroids mapped by stereographic projections and characterized by a potential energy surface built from a linear combination of decoupled double wells shaped purposely to create rare events over a range of temperatures. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4855675]

I. INTRODUCTION

The rare-event sampling problem in Monte Carlo methods[1–4] continues to provide extraordinary challenges, and at the same time, fertile ground for the development of clever simulation strategies. A recent significant contribution to a long list of such approaches[5–25] is the work of Plattner *et al.*[24] Like parallel tempering,[18–23] j-walking,[13] and other related methods, the Infinite Swapping[24] (INS) and Partial Infinite Swapping (PINS) strategies provide a systematic way of improving sampling in particularly challenging Monte Carlo simulations. The work of Plattner *et al.*[24] focuses on the stochastic estimation of a physical property f over a distribution $\pi(x, T_k)$ at temperature T_k, by the following configuration integral:

$$\langle f \rangle_k = \frac{\int \pi(x, T_k)\, f(x)\, dx}{\int \pi(x, T_k)\, dx}. \tag{1}$$

The sampling challenges for Monte Carlo methods can be gleaned from graphs of the distribution, $\pi(x, T_k)$, which at temperatures near phase changes and rare events, are sparse and have disconnected domains around potential minima.[24] INS and the related PINS strategies replace $\pi(x, T_k)$ with a symmetrized joint probability distribution for N temperatures and N independent systems with coordinates $\mathbf{x}_n$,

$$\langle f \rangle_k = \frac{1}{N!} \sum_m \frac{\int P_m \left[\pi(\mathbf{x}_1, T_1)\, \pi(\mathbf{x}_2, T_2) \cdots \pi(\mathbf{x}_{,N}\, T_N)\right] f\left(\mathbf{x}_{idnx(n,k)}\right) d\mathbf{X}}{\int P_n \left[\pi(\mathbf{x}_1, T_1)\, \pi(\mathbf{x}_2, T_2) \cdots \pi(\mathbf{x}_N, T_N)\right] d\mathbf{X}}, \tag{2}$$

where

$$d\mathbf{X} = d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N,$$

P_n is the permutation operator, and $idnx(n, k)$ is the pointer to all the permutations in the coordinate-temperature list associated with temperature k.[24] The symbol $\mathbf{x}_i$ is used to notate the n-dimensional set of coordinates $x^1, \ldots, x^n$ for the ith walker, and the superscripts on x are not powers, rather they are used to distinguish the degrees of freedom. The authors of Ref. 24 demonstrate that the symmetrized distribution in Eq. (2) is highly connected if the range of temperatures is selected carefully. Infinite swapping and the partial infinite swapping methods[24] have been subjected to rigorous tests using systems like the well known LJ_{38} cluster, and the authors have reported that partial infinite swapping can be made more efficient than parallel tempering with reasonable means. The

a)Electronic mail: curotto@arcadia.edu
b)Electronic mail: massimo.mella@uninsubria.it

development of INS and PINS[24] is particularly timely, given our own interests in simulating molecular clusters,[26–41] and the difficulties we have encountered in the process. Molecular clusters seem to have a much richer set of rare-event like thermodynamic behaviors compared to an atomic cluster of equivalent size.[42–68] Our experience with ammonia clusters, for example, suggests that as few as 11 ammonia molecules[27] can cause significant sampling difficulties and slow converging results in regions of temperature dominated by phase changes, even when using state of the art methods like parallel tempering.

In a recent investigation,[25] we have identified a set of multidimensional potential energy surfaces of sufficient complexity to make the convergence of parallel tempering significantly difficult and unfeasible for sufficiently large number of dimensions. At the same time, the model is sufficiently simple to allow for the computation of all the relevant thermodynamic properties by deterministic methods. Our aim was to develop strategies to improve the severe sampling problems we encounter in simulating ammonia clusters, and we succeed in identifying smart darting[5–7] as a method that can be implemented successfully in diffusion Monte Carlo simulations since the number of minima contributing to the ground state wave function may be relatively small and many complications[7] can be avoided. The potential we develop[25] in the course of that work is a linear combination of decoupled double wells [$(\mathrm{DDW})_n$]

$$V(x^1, \ldots, x^n) = \sum_{i=1}^{n} V_1(x^i), \tag{3}$$

where

$$V_1(x) = ax^4 + bx^3 + cx^2 + 1, \tag{4}$$

and a, b, c are parameters used to shape the function continuously from a symmetric double well with a barrier of 1, ($b = 0$, $c = -a$) to a family of potentials with a minimum at a higher energy relative to the other. The function V_1 is a member of a family of quartic double wells used by Frantz to test j-walking.[13] The $(\mathrm{DDW})_n$ family of potentials can be fully characterized for any given value of n by using deterministic methods. The classical thermodynamic properties for $(\mathrm{DDW})_n$, such as the average potential, $\langle V(x^1, \ldots, x^n)\rangle$ can be recast in terms of their monodimensional counterpart,

$$\langle V(x^1, \ldots, x^n)\rangle = n\langle V_1(x)\rangle, \tag{5}$$

$$\langle E(x^1, \ldots, x^n)\rangle = \frac{n}{2} k_B T + n \langle V_1(x)\rangle, \tag{6}$$

$$\frac{C_V}{k_B} = \frac{n}{2} + n\frac{\langle V_1^2(x)\rangle - \langle V_1(x)\rangle^2}{(k_B T)^2}. \tag{7}$$

One may compute the integrals such as the monodimensional partition function

$$Q^{(1)} = \int_{-\infty}^{\infty} dx \, \exp\left[-\beta V_1(x)\right], \tag{8}$$

and those for $\langle V_1^2(x)\rangle$, and $\langle V_1(x)\rangle$, by using standard quadrature techniques.

For a given value of n, there are 2^n minima in total, and $n2^{n-1}$ first order saddles. Therefore, the stochastic evaluation of thermodynamic properties for $(\mathrm{DDW})_n$, via Eq. (1) with parallel tempering, can quickly become prohibitive as the size of the system n increases. We have shown[25] that the set $(\mathrm{DDW})_n$ creates serious challenges for parallel tempering simulations when $n > 20$. For these systems we have been unable to converge any parallel tempering simulations with feasible walk lengths and using the best of our abilities to optimize all the parameters of the simulation including the choice of the temperature schedule.[25]

The prime objective of the present work is to extend the INS and PINS methods to curved spaces. The larger goal we hope to achieve eventually is to extend the present reach of classical and quantum Monte Carlo methods into the emerging frontier of condensed molecular matter. In this article we report on the initial steps taken along this path. Simulations of complex molecular systems must often employ holonomic constraints, especially when the methods include quantum corrections, hence the need to explore curved spaces in general. The rationale for testing INS and PINS for these systems lies in the indication that the PINS algorithms may perform better than parallel tempering for extremely challenging systems. The INS and PINS methods[24] seem to be most successful and easier to implement when the sampling of the symmetrized joint probability in Eq. (2) is performed by dynamic methods.[69–71] With a view of employing more flexible sampling approaches, the methods[72–74] (*vide infra*) used in this work are evolved from the generalization of SMC and based on the recognition that a step of the latter is equivalent to a resampling of the degrees of freedom conjugated momenta with the appropriate Boltzmann distribution followed by a single trajectory integration step (carried out with the velocity Verlet's method[75]). Given the ease with which the appropriate deterministic benchmarks can be generated, and the fact that the convergence of parallel tempering on these systems has been measured carefully in previous studies[25] it behooves us to carry out tests on the PINS methods to explore strategies, test numerical integrators and determine the set of parameters that creates the best convergence properties for this class of algorithms.

The rest of this article is organized as follows: In Sec. II we report the results of a number of tests performed on the $(\mathrm{DDW})_n$ family of potentials in flat spaces mapped by Cartesian coordinates. Since it is relatively simple to generate converged results with deterministic methods, we use flat spaces in our preliminary numerical work to test our codes and to optimize the details of the algorithms before tackling the more complicated problems in curved manifolds. We believe that this further characterization of INS and PINS in spaces mapped by Cartesian coordinates could prove useful to a much larger community of researchers. In Sec. III we develop a generic integrator for systems in curved spaces based on the variational approach. The numerical tests on curved spaces are reported in Sec. IV and our conclusions are in Sec. V.

II. INFINITE SWAPPING AND PARTIAL INFINITE SWAPPING IN $\mathbb{R}^n$

Infinite swapping and partial infinite swapping are relatively new developments, and have never been tested with the

surface in Eq. (3) in $\mathbb{R}^n$ mapped with Cartesian coordinates. To sample the symmetrized joint probability distribution in (2) one defines[24] the quantity

$$\mu(X) = \sum_m P_m \left[\pi(\mathbf{x}_1, T_1)\,\pi(\mathbf{x}_2, T_2)\cdots\pi(\mathbf{x}_N, T_N)\right], \quad (9)$$

then,

$$\int P_m \left[\pi(\mathbf{x}_1, T_1)\,\pi(\mathbf{x}_2, T_2)\cdots\pi(\mathbf{x}_N, T_N)\right] dX = \frac{1}{N!}\int \mu(X)\,dX, \quad (10)$$

since all the terms of the sum on the right are identical. Consequently, Eq. (2) becomes

$$\langle f \rangle_k = \frac{1}{N!}\frac{\int \mu(X)\sum_{m=1}^{N!}\rho_m(x)\,f(x_{idnx(n,k)})\,dX}{\int \mu(X)\,dX}, \quad (11)$$

where

$$\rho_m(X) = \frac{P_m\left[\pi(\mathbf{x}_1, T_1)\,\pi(\mathbf{x}_2, T_2)\cdots\pi(\mathbf{x}_N, T_N)\right]}{\mu(X)}. \quad (12)$$

The authors of Ref. 24 explain in detail the INS approach using three temperatures as example, and for the sake of completeness we briefly repeat the outline here. With three temperatures we compute a total of nine values of $\pi(\mathbf{x}_k, T_k)$, i.e., all the possible permutations of the coordinates associated with the three temperatures, $\pi(\mathbf{x}_1, T_1)$, $\pi(\mathbf{x}_1, T_2), \ldots \pi(\mathbf{x}_3, T_3)$. From these we obtain the six terms of the sum of Eq. (12) by first computing the products, $\pi(\mathbf{x}_1, T_1)\,\pi(\mathbf{x}_2, T_2)\,\pi(\mathbf{x}_3, T_3)$, and the remaining five possible permutations of the coordinates, leaving the temperatures in the same order, e.g., $\pi(\mathbf{x}_2, T_1)\,\pi(\mathbf{x}_1, T_2)\,\pi(\mathbf{x}_3, T_3)$, etc. Once these products are formed, $\mu(X)$ is computed with Eq. (9), and the weights ρ_n with Eq. (12). To move we select a particular coordinate-temperature association according to the values ρ_n, following the standard procedure to sample sums over arbitrary distributions.[4]

Once a particular coordinate-temperature association is chosen, moves are made at all the temperatures. These can be performed by the Metropolis algorithm,[1–4] or with other strategies such as the Smart Monte Carlo.[69–71] In this work, we opted for using the isothermal molecular dynamics (IMD) by Andersen,[72,73] which is an evolution of SMC and equivalent to hybrid Monte Carlo in the limit of no rejection.[74] Notice that the theoretical basis for the generalization of SMC to more than a single trajectory integration step ($k_{\max} > 1$) leading to IMD had already been hinted in Refs. 69–71, and it is based on the equivalence of the single step SMC indicated in Sec. I and the phase space volume conservation of dynamical trajectories. To implement the IMD method in this work, we closely follow the suggestion made in Ref. 73 to sample the configuration Boltzmann distribution for a given potential energy. In practice, we use the temperature to scale Gaussian random numbers so to generate new momenta distributed according to the correct Boltzmann distribution and that, combined with the current position of the walker, become the starting point for a classical trajectory propagated by $k_{\max}$ steps using Δt as the time increment. The integrator we use in flat spaces is the Candy-Rozmus algorithm, a fourth order method. We note that INS3 works quite well for the monodimensional case and that its performance is insensitive to the parameters of the integrator Δt and $k_{\max}$, provided the integrator produces a stable numerical solution.

Practical applications of INS and PINS require many more temperatures, and coding the general PINS needs some additional automation to generate the $n!$ possibilities when n temperatures are involved. To this end, we formulate an algorithm to create a matrix representation of the permutation operators P_k of a list of k objects, using the following steps:

1. Initialize $N!$ square matrices, $N \times N$ is size: $\Gamma_l, l = 1, 2, \ldots, N!$. The first $N(N-1)/2 + 1$ of these are filled by placing a 1 along all the diagonal entries and zero everywhere else. The first one, Γ_1, is the identity operator.
2. For all values of $l = 2$ to $N(N-1)/2 + 1$, replace the i, i and j, j entries in Γ_l with zeros, and the i, j and j, i entries with one for $i = 1$ to k and for $j = i + 1$ to k.
3. Compute the product of all the distinct pairs of all the matrices obtained in the previous step.
4. Augment the set of distinct operators and repeat step 3 until all $k!$ operators have a matrix representation.

For the $N = 3$ example, a total of six permutation operators are formed. With $\Gamma_1(123) = (123)$, as the identity, and with $\Gamma_2(123) = (213)$, $\Gamma_3(123) = (321)$, and $\Gamma_4(123) = (132)$, as obtained in the second step, one can construct the following multiplication table:

	Γ_1	Γ_2	Γ_3	Γ_4
Γ_1	Γ_1	Γ_2	Γ_3	Γ_4
Γ_2	Γ_2	Γ_1	Γ_5	Γ_6
Γ_3	Γ_3	Γ_6	Γ_1	Γ_5
Γ_4	Γ_4	Γ_5	Γ_6	Γ_1,

(13)

where $\Gamma_5(123) = (231)$ and $\Gamma_6(123) = (312)$ are the remaining distinct permutation operators that consist of two consecutive single permutations obtained with a single iteration of step 3. The operators Γ_l generated by the algorithm presented earlier are used to find the set of $N!$ vectors with entries permuted by the set Γ_l to compute $\mu(X)$ with Eq. (9) and the weights ρ_n with Eq. (12) for the general case. Furthermore, the coordinate index associated with the averages at temperature T_j in Eq. (11) is computed from the permutation operators as well. For a given temperature T_j, there are $(N-1)!$ permutations that associate x_1 with T_j, $(N-1)!$ permutations that associate x_2 with T_j, etc. The averages at temperature T_j in Eq. (11) are the weighted average of all $N!$ permutations associated as described. These associations are used to construct the *idnx* array for the general case. For the $N = 3$ example and the Γ_{1-6} as defined earlier, at T_1, the walker array pointer is $idnx(n, 1) = 1, 2, 3, 1, 2, 3$, at T_2, it is $idnx(n, 2) = 2, 1, 2, 3, 3, 1$ and so on.

In Fig. 1, the heat capacity of several INS and PINS simulations is compared for the $n = 2$ case. The continuous lines in the heat capacity are obtained with Eqs. (5)–(7). The INS6 simulation in Fig. 1 (green circles) uses the following values of k_BT: 0.01, 0.03, 0.06, 0.09, 0.12, and 0.15 hartree. The

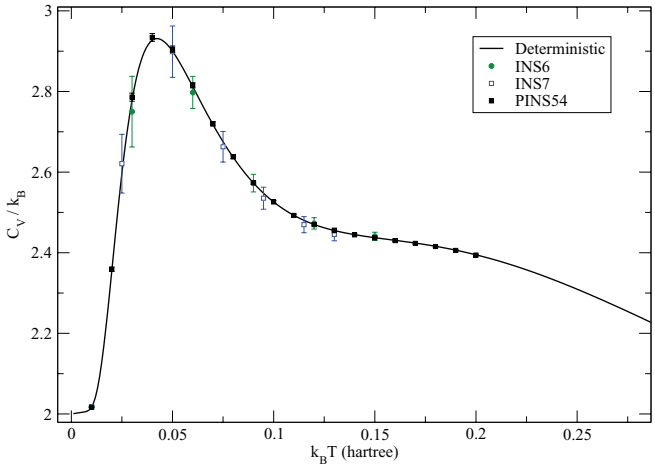

FIG. 1. Classical NVT heat capacity for a particle in $\mathbb{R}^2$, with 1 a.u. of mass, subjected to the double well in Eq. (3) at several values of $k_B T$.

INS7 simulation in Fig. 1, uses 0.01, 0.025, 0.050, 0.075, 0.095, 0.115, and 0.13 hartree as the values of $k_B T$. Both INS6 and INS7 results are obtained by averaging 15 independent blocks of 10^7 moves, each move being composed of a resampling of the momenta and a trajectory integration using the fourth order Candy-Rozmus integrator with $\Delta t = 0.01$ a.u. and $k_{\max} = 10$. All error bars in Fig. 1 represent a 95% confidence interval computed with the appropriate student-t distribution value.

A simulation with $n > 2$ requires more than 10 temperatures, and an infinite swapping simulation requires a symmetrized joint distribution with more than 3.6×10^6 permutations for $N = 10$. To handle the factorial growth, the authors[24] of the infinite swapping techniques have developed a "dual chain" process by which a N-temperature set is partitioned in two distinct ways. Each chain of temperatures is divided into blocks of relatively small size, to reduce the local shuffling to a manageable task. This is accomplished by symmetrizing only within the blocks of each chain. By carefully selecting the two partitions so as to not leave common gaps between temperature blocks, and by exchanging configurations from one chain to the other after every move, PINS generates the fully symmetrized distribution in Eq. (11) asymptotically.[24] We experiment with several possibilities, including a 25 temperatures PINS(1-3|3-1) scheme. The notation PINS(1-3|3-1) in conjunction with the number of temperatures indicates that the first chain of temperatures contains one block with one temperature and eight consecutive blocks with three temperatures, while the second chain contains eight consecutive blocks with three temperatures followed by one block with the highest temperature. Therefore, the two chains do not have a common temperature gap between the blocks. The 25 temperatures PINS(1-3|3-1) scheme succeeds for relatively small size systems (among which the $(\mathrm{DDW})_2$ in Fig. 1). Increasing the dimensionality of the system simulated, however, causes the convergence of PINS(1-3|3-1) to slow down considerably, and with as few as 10 dimensions parallel tempering yields better results with 10^6 potential points per temperature, compared to the equivalent PINS(1-3|3-1).

To improve on the situation described above, a new set of temperature schedules is developed and tested. The PINS54 data in Fig. 1 (and also *vide infra* Figure 2) are from 20 temperature simulations created with a dual chain, one composed of consecutive blocks of five temperature, the other with consecutive blocks of four temperatures. The first common temperature gap in a general PINS*nm* scheme occurs at the lowest common multiple of n and m. The first common temperature gap occurs after 20 temperatures for the PINS54 case. Notice how the standard error associated to the PINS54 data is much reduced compared to INS ones as it is for PINS(1-3|3-1).

The reduction of the standard error associated to the heat capacity when a properly chosen temperature schedule is employed does not, however, necessarily indicate that PINS is more efficient than other methods, like the Replica Exchange (RE), for example. To gauge the possible gains in efficiency, we compare quantitatively the convergence properties of PINS54 and replica exchange (both using 10 steps for every move) in $\mathbb{R}^{30}$, a situation in which parallel tempering is already known to fail.[25] We do this by computing the ergodic measure[77] for several values of the number of dynamic segments k each comprised of $k_{\max}$ steps,

$$d_k = \frac{2}{M(M-1)} \sum_{i=2}^{M} \sum_{j=1}^{i-1} \left[\overline{V}_k^{(i)} - \overline{V}_k^{(j)} \right]^2, \tag{14}$$

with M being the number of independent random walks performed. d_k is the mean square deviation of the potential energy average computed with k steps:

$$\overline{V}_k^{(i)} = \frac{1}{k} \sum_{j=1}^{k} V\left(x_k^{(j)}\right), \tag{15}$$

and $x_k^{(j)}$ is the configuration after j segments in walk i. Here, replica exchange is a procedure equivalent to parallel tempering with respect to temperature swaps, but it differs in the fact that the exploration of the configurational space is carried out using trajectories of a predefined number of steps rather than random displacements. When the integration is carried out with only one step ($k_{\max} = 1$) after a resampling of the momenta, the procedure is strictly equivalent to a SMC simulation for each temperature despite the fact that no rejection check is carried out in RE as the Δt is chosen small enough to guarantee energy conservation. Notice that SMC is usually considered able to explore more efficiently the configuration space than standard Monte Carlo with random displacements.[69,70]

The graphs in the top panel of Fig. 2 display the expected behavior[21]

$$d_k = A_k \frac{1}{k} + B_k + \gamma_k, \tag{16}$$

where A_k and γ_k are the short time constant and the noise, respectively, whereas B_k is a slowly decreasing function of k. As a simulation converges, the first two terms on the right-hand side of Eq. (16) should drop below the noise. The two sets of simulations compared in Fig. 2 clearly demonstrate that the short time constant is comparable among them. However, the long time contribution to the ergodic measure for parallel

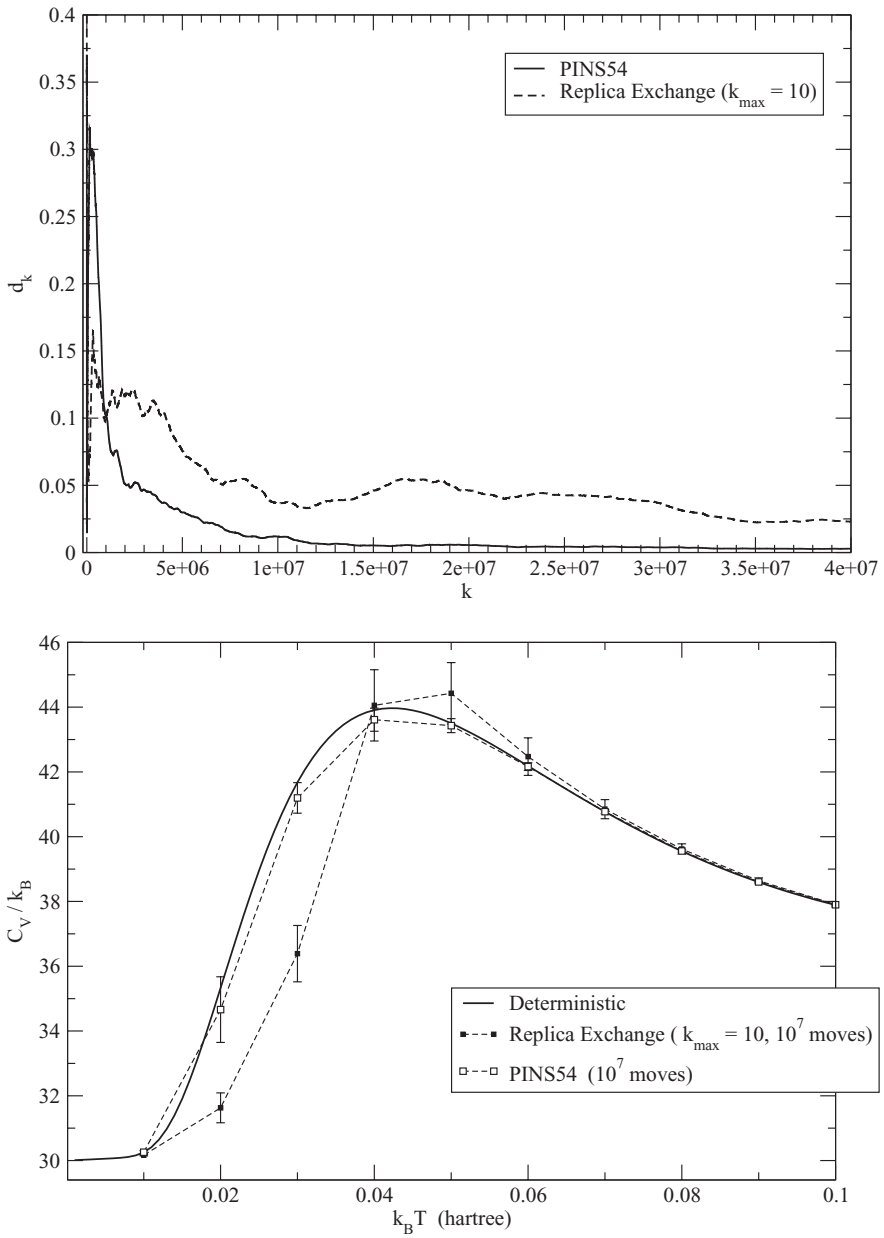

FIG. 2. Top: Ergodic measure for PINS54 and replica exchange (units of hartree^2) for $(DDW)_{30}$ as a function of the number of dynamic segments k. Bottom: classical NVT heat capacity for a particle in $\mathbb{R}^{30}$, with 1 a.u. of mass, subjected to the double well in Eq. (3) at several values of k_BT. A replica exchange and a partial infinite swapping simulations requiring the same amount of time are compared. Both sets of results are obtained by running 22 independent samples each with 10^7 trajectories with $k_{max} = 10$, after a "warm up run" of equal length.

tempering (B_k) is at least one order of magnitude greater than the same quantity for PINS54.

The differences in B_k and γ_k between PINS54 and RE, evident from the previous tests may have important consequences when computing observables that are sensitive to the sampling ergodicity. In the bottom panel of Fig. 2 we compare the heat capacity computed for $(DDW)_{30}$ using RE[76] and PINS54 simulations with the results expected from the deterministic computation in Eq. (7). In this case, the replica exchange simulation results are obtained by running 22 independent samples each with 10^7 $k_{max} = 10$ segments, after a "warm up run" of equal length. This simulation is approximately equally expansive as the PINS54 simulations with 10^7 moves and a $k_{\max} = 10$. The error bars represent the 95% confidence interval and are obtained from the averages of each sample. The PINS54 results are created by averaging 22 independent samples each consisting of 10^7 moves, after a "warm up run" of equal length. The IMD trajectories used in both RE and to sample the PINS54 distribution are propagated by the fourth order Candy-Rozmus integrator with $\Delta t = 0.01$ a.u. as in Fig. 1.

The data in the bottom panel of Fig. 2 show that PINS54 is more accurate than replica exchange for the same number of moves. The data points for the latter are significantly and systematically below the deterministic results between the melting peak and the equipartition limit. Thus, while PINS(1-3|3-1) performs worse than the PT simulation when $N > 10$ as indicated, PINS can be systematically improved,

and PINS54 seems already to be better than the RE simulations in Fig. 2.

Additional computational experiments were carried out to gain a better understanding of the performance obtainable while using PINS54. Thus, we found that the proper walk length for PINS54 to converge results for $(DDW)_{40}$ is considerably longer: based on the data trends at the lowest temperatures we estimate that as many as 3×10^9 moves may be needed to obtain comparable accuracy to the $(DDW)_{30}$ case. We also discovered that the success shown by PINS54 in Fig. 2 depends on optimizing k_{max}, and Δt. A calculation with $k_{max} = 1$, for example, does not converge any better on $\mathbb{R}^{30}$ than replica exchange, while it does so when $k_{max} = 10$ (see Fig. 2). We learned that k_{max} has to be sufficiently large to allow those trajectories that start with sufficiently large momentum to overcome the barriers. A Δt ten times smaller would require a k_{max} ten times larger to create the same opportunities. We did not systematically characterize the behavior of PINS54 with sets of Δt and k_{max}. Instead, we choose as large a Δt as possible, that provides a reasonable energy conservation, and determine the smallest value of k_{max} needed to travel from a minimum to the farthest transition state that leads to another minimum of statistical importance. Above $n = 20$, for each dimension, there is a minimum walk length required to reach convergence regardless of the value chosen for k_{max} and Δt. A PINS54 simulation with $k_{max} = 10$ and $\Delta t = 0.1$ a.u. in $\mathbb{R}^{40}$ is statistically identical and equally off the mark, as a PINS54 simulation with $k_{max} = 10$ and $\Delta t = 0.01$ a.u. with equal number of moves, for example.[78]

III. A VARIATIONAL INTEGRATOR FOR CURVED SPACES

The results in flat spaces encourage us to continue in our goal to extend the PINS methods to curved spaces. From the simulations in flat spaces we learn that IMD with a number of steps is necessary to converge asymptotically with reasonable means, for extremely challenging rare events. We learn that energy conservation is crucial for the optimal functioning of the PINS algorithm. There are methods for integrating Euler-Lagrange's equations for systems subjected to holonomic constraints. The SHAKE[79] and RATTLE[80] algorithms are two popular examples. These are variational algorithms derived by using augmented Lagrangian functions to introduce the constraints. In this paper, we are interested in more general curved spaces beyond those generated by holonomic constraints,[81,82] as these may arise in path integral simulations, for example.[83,84] In a recent application of Ring Polymer Dynamics to curved spaces,[84] we develop a method to integrate directly from generalized coordinates, and we observe that direct application of traditional symplectic integrators gives rise to numerical methods with significant amount of energy drift. The method we have developed in Ref. 84 to improve the long term energy drift of the symplectic integrator is to compute the state vector by using the Liouville operator. The latter is not split as customary. Rather, to accommodate the momenta-coordinate coupling that takes place in general curved spaces, we create a Krylov space using a matrix representation of the Liouville operator. Let $\mathcal{H}$ represent the Hamiltonian of the system, $\mathbf{q}$ the set of generalized coordinates, and $\mathbf{p}$ the corresponding canonical conjugate momenta. Then, the phase space vector is updated with

$$\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix}_{t+\Delta t} = \begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix}_{t} + \Delta t \begin{pmatrix}\dfrac{\partial \mathcal{H}}{\partial \mathbf{p}}\\ -\dfrac{\partial \mathcal{H}}{\partial \mathbf{q}}\end{pmatrix} + \frac{1}{2}\Delta t^2 \begin{pmatrix}-\dfrac{\partial \mathcal{H}}{\partial \mathbf{p}}\dfrac{\partial^2 \mathcal{H}}{\partial \mathbf{p}\partial \mathbf{q}}\\ \dfrac{\partial \mathcal{H}}{\partial \mathbf{q}}\dfrac{\partial^2 \mathcal{H}}{\partial \mathbf{q}\partial \mathbf{p}}\end{pmatrix} + \ldots + \mathcal{O}(\Delta t^3). \quad (17)$$

The first two terms of Eq. (17) can be derived with the split operator approach. The last term in Eq. (17) vanishes in $\mathbb{R}^d$, or in spaces endowed with a uniform metric, and would be absent even when the mixed partial derivatives are not zero if we had split the Liouville operator. This approach was tested by developing a series of algorithms that converge to higher and higher order. However, we have noted that the energy conservation property $(E - E_0)/E_0$, where E_0 is the energy at $t = 0$ always scales as Δt, and that there is always a measurable amount of drifting regardless of the order of integration. Yet, the class of Hamiltonians we use is invariant under time translation, therefore the energy should be strictly conserved by a "good" algorithm, and this property is desired for the present application. Therefore, we are pushed to develop new alternatives. The approach we develop below is a generalization of the variational method used to derive RATTLE and SHAKE, but it makes no use of Lagrangian multipliers, and integrates directly the phase space state expressed with the generalized coordinates and momenta. Integrators derived directly from Hamilton's principle are automatically endowed with a number of desired properties.[82] In particular, they preserve momenta associated to symmetry elements of the system, display excellent longtime energy conservation when the Hamiltonian is time invariant, and are symplectic in nature. Additionally, the variational methodology allows one to derive good integrators for systems characterized by extremely complex geometries[81,82] as we have in the present case.

We begin with the discrete Euler-Lagrange equations,

$$D_2 L_d\left(q_{k-1}, q_k, \Delta t\right) + D_1 L_d\left(q_k, q_{k+1}, \Delta t\right) = 0. \quad (18)$$

These are central to the implementation of variational algorithms, and are derived by approximating the variation of the classical action integral

$$\delta S = \delta \int \mathcal{L}\left(q\left(t\right), \dot{q}\left(t\right)\right) dt = 0, \quad (19)$$

with a quadrature. D_i represents the derivative of the discretized Lagrangian L_d, with respect to the ith argument, and the notation q_k is used to represent the entire set of coordinates q^μ $\mu = 1, 2, \ldots n$ at a finite value of time $t = t_k$. In this article, Greek superscripts are used as contravariant labels for coordinates, and for tensor quantities that transform contravariantly under a change of coordinates rather than representing powers. Greek subscripts are used as covariant labels for

conjugate momenta or n-forms.[81] The discretized Lagrangian L_d is obtained by approximating the velocity space $\dot{q}$ with a finite difference, and it represents the quadrature estimate of the integral in Eq. (19). Even ordered algorithms that propagate trajectories from the initial momentum and position (p_0, q_0) can be derived by using the discrete version of the momentum as defined from the derivatives of L_d,

$$p_k = -D_1 L_d\left(q_k, q_{k+1}, \Delta t\right), \tag{20}$$

$$p_{k+1} = D_2 L_d\left(q_k, q_{k+1}, \Delta t\right). \tag{21}$$

Setting $k \to k-1$ in Eq. (21) and substituting the result in Eq. (18) yields Eq. (20). This proves that both the discrete Euler-Lagrange equations and the discrete version of the momentum p_k and p_{k+1} are self consistent. Generally, one derives a variational-type integration algorithm by solving for the set q_{k+1} from Eq. (20) and uses the values of q_{k+1} to compute the set p_{k+1} with Eq. (21).

We seek to produce a second order integrator in position momentum form using the following general Lagrangian for systems in curved spaces,

$$\mathcal{L}\left(q, \dot{q}, t\right) = \frac{1}{2} g_{\mu\nu}\, \dot{q}^{\mu} \dot{q}^{\nu} - V, \tag{22}$$

where the 2-form $g_{\mu\nu}$ is the metric tensor of the space. In Eq. (22), we are making use of Einstein's shorthand for summation over repeated Greek indices in the upper and lower position. The kinetic energy term in this, contains an implied double sum over μ and ν from 1 to the dimension of the space n. The metric tensor $g_{\mu\nu}$ becomes a diagonal matrix with the mass as the element in flat spaces mapped by Cartesian coordinates. In curved spaces, the metric tensor typically depends on the generalized coordinates $q^{\mu}\mu = 1, 2, \ldots n$ coupling the coordinates and the momenta as a consequence. When we apply the trapezoid rule to the integral in Eq. (19) and use Eq. (22) we obtain the following expression:

$$L_d\left(q_k, q_{k+1}, \Delta t\right) = \frac{\Delta t}{2} \mathcal{L}\left(q_k, \frac{q_{k+1} - q_k}{\Delta t}\right) + \frac{\Delta t}{2} \mathcal{L}\left(q_{k+1}, \frac{q_{k+1} - q_k}{\Delta t}\right). \tag{23}$$

The elements inside the parenthesis ( ) are the arguments of the functions L_d and $\mathcal{L}$. We introduce the following shorthand $g_{\mu\nu k} = g_{\mu\nu}(q_k)$ and $V_k = V\left(q_k\right)$ to represent the metric tensor elements and the potential energy, respectively, evaluated at a finite value of time $t = t_k$. The metric tensor and the potential do not need to depend explicitly on time, but they do depend on it implicitly via the dependence of the coordinates q^{μ}. Inserting Eq. (22) into Eq. (23) gives

$$\begin{aligned} L_d\left(q_k, q_{k+1}, \Delta t\right) &= \frac{\Delta t}{4} g_{\mu\nu k} \left(\frac{q^{\mu}_{k+1} - q^{\mu}_k}{\Delta t}\right)\left(\frac{q^{\nu}_{k+1} - q^{\nu}_k}{\Delta t}\right) - \frac{\Delta t}{2} V_k \\ &+ \frac{\Delta t}{4} g_{\mu\nu k+1} \left(\frac{q^{\mu}_{k+1} - q^{\mu}_k}{\Delta t}\right)\left(\frac{q^{\nu}_{k+1} - q^{\nu}_k}{\Delta t}\right) - \frac{\Delta t}{2} V_{k+1}. \end{aligned}$$

Equations (20) and (21) become

$$\begin{aligned} p_{\mu k} &= \frac{1}{2} g_{\mu\nu k} \left(\frac{q^{\nu}_{k+1} - q^{\nu}_k}{\Delta t}\right) + \frac{1}{2} g_{\mu\nu k+1} \left(\frac{q^{\nu}_{k+1} - q^{\nu}_k}{\Delta t}\right) \\ &- \frac{\Delta t}{4} \partial_{\mu}\, g_{\sigma\nu k} \left(\frac{q^{\sigma}_{k+1} - q^{\sigma}_k}{\Delta t}\right)\left(\frac{q^{\nu}_{k+1} - q^{\nu}_k}{\Delta t}\right) + \frac{\Delta t}{2} \partial_{\mu} V_k \end{aligned} \tag{24}$$

and

$$\begin{aligned} p_{\mu k+1} &= \frac{1}{2} g_{\mu\nu k} \left(\frac{q^{\nu}_{k+1} - q^{\nu}_k}{\Delta t}\right) + \frac{1}{2} g_{\mu\nu k+1} \left(\frac{q^{\nu}_{k+1} - q^{\nu}_k}{\Delta t}\right) \\ &+ \frac{\Delta t}{4} \partial_{\mu}\, g_{\sigma\nu k+1} \left(\frac{q^{\sigma}_{k+1} - q^{\sigma}_k}{\Delta t}\right)\left(\frac{q^{\nu}_{k+1} - q^{\nu}_k}{\Delta t}\right) \\ &- \frac{\Delta t}{2} \partial_{\mu}\, V_{k+1}. \end{aligned} \tag{25}$$

Equation (24) represents a set of n independent coupled nonlinear algebraic equations. The root q^{μ}_{k+1} depends, in general, on the values of all the other roots, and the solution for such nonlinear coupled system can only be found iteratively. A good starting point for q^{μ}_{k+1} can be supplied by Eq. (17). For flat spaces mapped by Cartesian coordinates, $x^{\mu} \in \mathbb{R}^n$ the metric tensor is simply the mass associated with the ith degree of freedom,

$$g_{\mu\nu k} \to m\, \delta_{\mu\nu}, \tag{26}$$

and Eqs. (20) and (21) simplify to

$$p_{\mu k} = m\, \delta_{\mu\nu} \left(\frac{x^{\nu}_{k+1} - x^{\nu}_k}{\Delta t}\right) + \frac{\Delta t}{2} \frac{\partial V_k}{\partial x^{\mu}_k} \tag{27}$$

and

$$p_{\mu k+1} = m\, \delta_{\mu\nu} \left(\frac{x^{\nu}_{k+1} - x^{\nu}_k}{\Delta t}\right) - \frac{\Delta t}{2} \frac{\partial V_{k+1}}{\partial x^{\mu}_{k+1}}. \tag{28}$$

The resulting method is the position-momentum version of the variational Verlet algorithm[85] and its convergence displays second order behavior in Δt.

IV. NUMERICAL TESTS IN $\mathbb{T}^n$

In the following, we present a set of tests carried out on the new variational integrator scheme to confirm its performances. Also, the new integrator would be applied to carry out the trajectory evolution needed in IMD and for the appropriate sampling in PINS simulations.

A. Variational integrator

For a particle in a ring ($x^2 + y^2 = R^2$), mapped with a stereographic projection ξ, defined by the map $\Phi : \mathbb{T} \to \mathbb{R}^2$,

$$x = \frac{4R^2 \xi}{\xi^2 + 4R^2}, \qquad y = \frac{\xi^2 - 4R^2}{\xi^2 + 4R^2} R, \tag{29}$$

the metric tensor has a single element,

$$g = \frac{\left(4R^2\right)^2}{(\xi^2 + 4R^2)^2}, \tag{30}$$

and Eq. (22) reads

$$\mathcal{L}\left(\xi\left(t\right),\dot{\xi}\left(t\right)\right)=\frac{8mR^4}{(\xi^2+4R^2)^2}\dot{\xi}^2-V\left(\xi\right). \tag{31}$$

The conjugate momentum p is

$$p=\frac{\partial\mathcal{L}}{\partial\dot{\xi}}=\frac{16mR^4}{(\xi^2+4R^2)^2}\dot{\xi}. \tag{32}$$

When we apply the trapezoid rule, the discretized Lagrangian is

$$\begin{aligned}&L_d\left(\xi_k,\xi_{k+1},\Delta t\right)\\&=\frac{\Delta t}{2}\frac{8mR^4}{\left(\xi_k^2+4R^2\right)^2}\left(\frac{\xi_{k+1}-\xi_k}{\Delta t}\right)^2-\frac{\Delta t}{2}V\left(\xi_k\right)\\&\quad+\frac{\Delta t}{2}\frac{8mR^4}{\left(\xi_{k+1}^2+4R^2\right)^2}\left(\frac{\xi_{k+1}-\xi_k}{\Delta t}\right)^2-\frac{\Delta t}{2}V\left(\xi_{k+1}\right),\end{aligned} \tag{33}$$

and, Eqs. (20) and (21) become

$$\begin{aligned}p_k&=8mR^4\left(\frac{\xi_{k+1}-\xi_k}{\Delta t}\right)\left[\frac{1}{\left(\xi_{k+1}^2+4R^2\right)^2}+\frac{1}{\left(\xi_k^2+4R^2\right)^2}\right]\\&\quad+\frac{16mR^4\xi_k}{\left(\xi_k^2+4R^2\right)^3}\frac{(\xi_{k+1}-\xi_k)^2}{\Delta t}+\frac{\Delta t}{2}\frac{dV\left(\xi_k\right)}{d\xi_k},\end{aligned} \tag{34}$$

$$\begin{aligned}p_{k+1}&=8mR^4\left(\frac{\xi_{k+1}-\xi_k}{\Delta t}\right)\left[\frac{1}{\left(\xi_{k+1}^2+4R^2\right)^2}+\frac{1}{\left(\xi_k^2+4R^2\right)^2}\right]\\&\quad-\frac{16mR^4\xi_{k+1}}{\left(\xi_{k+1}^2+4R^2\right)^3}\frac{(\xi_{k+1}-\xi_k)^2}{\Delta t}-\frac{\Delta t}{2}\frac{dV\left(\xi_{k+1}\right)}{d\xi_{k+1}}.\end{aligned} \tag{35}$$

The potential used to test the method is in Eq. (4) with $x_i=\xi$, and with the parameters $a=1.02651$, $b=-0.05302$, and $c=-1.97349$. The units of these parameters are hartree bohr^{-4} for a, hartree bohr^{-3} for b, etc. The graph of the quartic polynomial in Eq. (4) with these choices has an asymmetric double well feature with the lowest minimum $V_1=0$ at $\xi=1$, a higher energy minimum $V_1=0.1$ at $\xi=-0.96126$ bohr, and a barrier of 1 hartree at $\xi=0$.

We test the algorithm using a particle of unit mass in a ring with a radius of 0.7 bohr. The graph in Fig. 3(a) is the phase space profile of a single trajectory started at $\xi=1.03634$ bohr, $p=0.684556$ a.u., and propagated for 10^5 consecutive segments using a $\Delta t=0.01$ a.u. Bracketing is used to solve for ξ_{k+1} from Eq. (34). The bracketing procedure is iterated 40 times reducing the uncertainty in the root, typically, to less than one part in 10^{12}. We find that this amount of precision is necessary to achieve the proper stability (i.e., lack of energy drift), and we adopt this procedure in all consecutive tests. The graph in Fig. 3(b) is the relative energy difference as a function of time, $(E-E_0)/E_0$, where E is computed with

$$E=\frac{\left(\xi^2+4R^2\right)^2}{8mR^4}p^2+V\left(\xi\right), \tag{36}$$

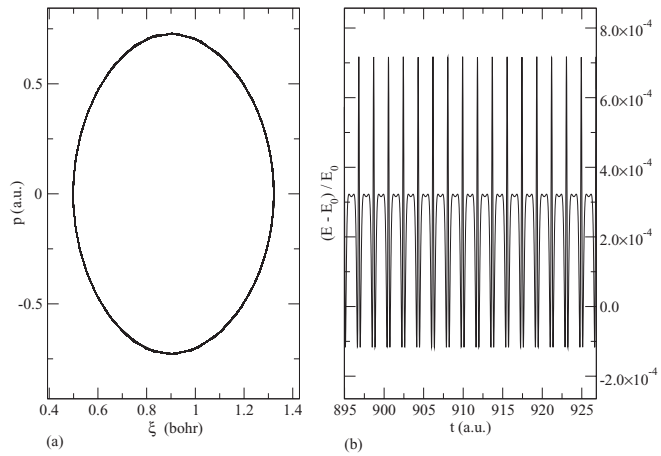

FIG. 3. Single trajectory for a particle in a ring, 1 a.u. of mass, and a radius of 0.7 bohr mapped with a stereographic projection coordinate. The potential energy is the double well in Eq. (4) and $\Delta t=0.01$ a.u. (a) Phase space projection. (b) Relative energy difference as a function of time.

and E_0 is the energy at $t=0$. We repeat these calculations using higher energies where the barrier crossing takes place. The results are in Fig. 4. The phase space trajectory is shaped with the expected double feature, while the $(E-E_0)/E_0$ graph looks nearly identical to the one in Fig. 3(b). By running a 10^6 step trajectory with $\Delta t=0.001$ a.u., [red line in Fig. 4(b)] we observe a factor of one hundred reduction in amplitude of the relative energy difference over time. Therefore, the algorithm converges quadratically both on the trajectory and on the correct energy and it represents a substantial improvement over the previous methods we have proposed for direct integrations in curved spaces using non-Cartesian coordinates. In all cases, we find no evidence of energy drift over time. The iterative nature of the method we develop here does add some computational overhead compared to a second order integrator in flat spaces. However, the present method does not require any additional potential (or derivative) evaluations when compared with its equivalent integrator in flat spaces, and therefore, for

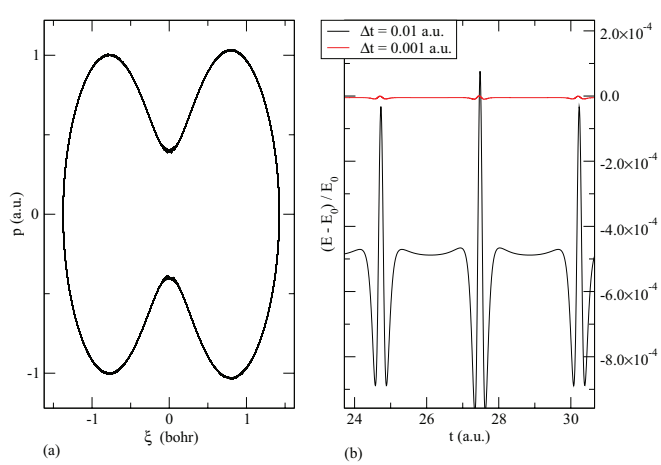

FIG. 4. Single trajectory for a particle in a ring, 1 a.u. of mass, and a radius of 0.7 bohr mapped with a stereographic projection coordinate. The potential energy is the double well in Eq. (4) and the energy is greater than the barrier height. (a) Phase space projection. (b) Relative energy difference as a function of time with $\Delta t=0.01$ a.u. for the black line, and $\Delta t=0.001$ a.u. for the red line.

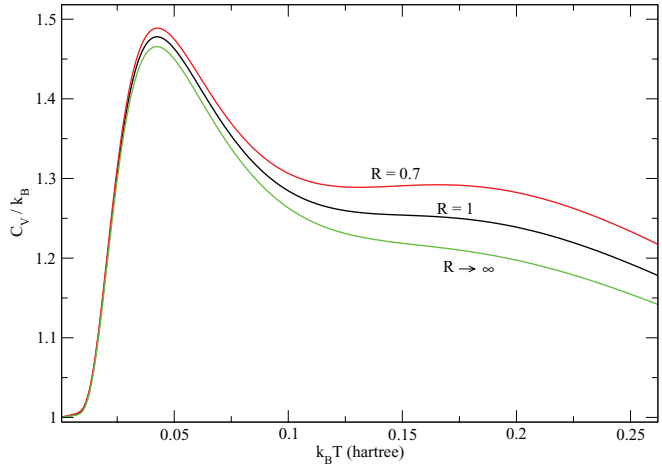

FIG. 5. Classical NVT heat capacity in units of the Boltzmann constant k_B, for a particle in a ring, 1 a.u. of mass mapped with a stereographic projection coordinate ξ, for various values of the radius R.

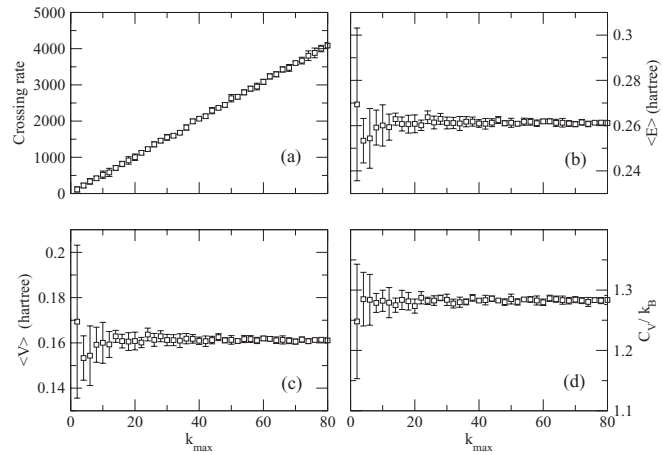

FIG. 6. Classical NVT ensemble averages for a particle in a ring of radius 0.7 bohr, 1 a.u. of mass mapped with a stereographic projection coordinate for various values of $k_{\max}$, the total number of dynamic steps. The potential energy is the double well in Eq. (4). The time steps and the temperature are $\Delta t = 0.01$ a.u., $k_B T = 0.2$ hartree, respectively. (a) The number of trajectories from the 10^6 ensemble that cross the barrier exactly once. (b) Canonical average energy in hartree. (c) Canonical average potential energy in hartree. (d) Heat capacity in units of the Boltzmann constant k_B.

expensive potential energy models, the added computational cost can be a minimal fraction of the whole.

B. IMD

The next set of numerical tests consists of computing the NVT ensemble properties for a particle of unit mass in a ring with a radius of 0.7 bohr for several values of $k_B T$ between 0.02 hartree and 0.4 hartree. The family of monodimensional configuration integrals,

$$\langle f \rangle_1 = \frac{\int_{-\infty}^{\infty} d\xi \, \left(\xi^2 + 4R^2\right)^{-1} \exp\left(-\beta V\right) \, f}{\int_{-\infty}^{\infty} d\xi \, \left(\xi^2 + 4R^2\right)^{-1} \exp\left(-\beta V\right)}, \tag{37}$$

with $\beta = (k_B T)^{-1}$, can be relatively easily computed with the trapezoid rule. From these it is simple to construct the canonical average energy $\langle E \rangle_n = n \langle E \rangle_1$ for the monodimensional case.

In Fig. 5 we graph the deterministic value of $C_V = (k_B T)^{-2}(\langle E^2 \rangle_1 - \langle E \rangle_1^2)$, in units of the Boltzmann constant k_B, for a particle in a ring, 1 a.u. of mass mapped with a stereographic projection coordinate ξ, for various values of the radius R for the monodimensional case. The graph of this quantity suggests that the impact of the topology on the thermodynamic properties can be substantial. As $R \to \infty$, $g \to 1$, the manifold becomes a flat space, and the results in Ref. 25 are reproduced exactly. We use the deterministic results to test the IMD approach in curved spaces.[72,73]

To compute the integral over the phase space

$$\langle f \rangle_1 = \frac{\int_{-\infty}^{\infty} dp \, d\xi \, \exp\left(-\beta \mathcal{H}\left(\xi, p\right)\right) \, f}{\int_{-\infty}^{\infty} dp \, d\xi \, \exp\left(-\beta \mathcal{H}\left(\xi, p\right)\right)}, \tag{38}$$

stochastically, with the Hamiltonian given as

$$\mathcal{H}\left(\xi\left(t\right), p\left(t\right)\right) = \frac{\left(\xi^2 + 4R^2\right)^2}{8mR^4} \, p^2 + V\left(\xi\right), \tag{39}$$

we run 10^6 trajectories, evaluating for each the following averages:

$$\overline{f}_i = \frac{1}{\tau} \int_0^{\tau} dt f\left(\xi\left(t\right), p\left(t\right)\right), \quad \xi\left(t=0\right), p\left(t=0\right) = \xi_i, p_i, \tag{40}$$

$$\langle f \rangle = \frac{1}{M} \sum_{i=1}^{M} \overline{f}_i. \tag{41}$$

The variational integrator developed here is used to propagate each trajectory from the final position ξ_i of the previous trajectory, and a randomly selected initial value of the momentum $p = p_i$ for a number of steps $k_{\max}$, therefore $\tau = k_{\max} \Delta t$. For this step, a Gaussian distribution with a standard deviation σ given by

$$\sigma = \frac{4mR^2}{\left(\xi_i^2 + 4R^2\right)} \sqrt{k_B T}, \tag{42}$$

is sampled using the Box-Muller algorithm.[4]

To demonstrate that the procedure indicated above converges to the correct results, in Fig. 6 we graph the Classical NVT ensemble averages for a particle in a ring of radius 0.7 bohr, 1 a.u. of mass mapped with ξ for several values of $k_{\max}$, $\Delta t = 0.01$ a.u., and $k_B T = 0.2$ hartree. The reported values are the mean of ten block averages with each block consisting of 10^6 trajectories. The error bars are the standard deviations computed from the individual block averages. At $k_{\max} = 80$, $k_B T = 0.2$ hartree, $\langle E \rangle$ is 0.2612 ± 0.0011 hartree, $\langle V \rangle$ is 0.1611 ± 0.0011 hartree, and C_V / k_B is 1.2837 ± 0.0076. These values are in excellent agreement with the deterministic ones, 0.26133 hartree, 0.16133 hartree, and 1.28247, respectively.

Providing a more thorough test, the graphs in Fig. 7 are for the classical NVT heat capacity at several values of $k_B T$, for a particle in a twenty-dimensional toroid, $\mathbb{T}^{20}$, with 1 a.u. of mass, a radius uniform in all dimensions and equal to

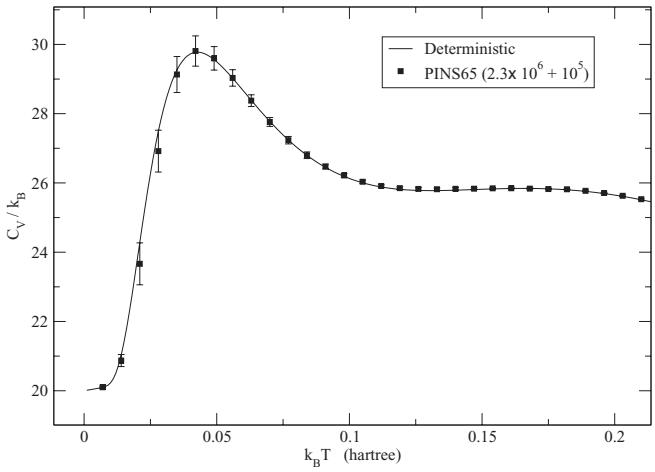

FIG. 7. Simulated heat capacity for a particle in $\mathbb{T}^{20}$, with 1 a.u. of mass, a radius uniform in all dimensions equal to 0.7 bohr, and subjected to the double well in Eq. (3).

0.7 bohr mapped with stereographic coordinates. The space is characterized by the double well model in Eq. (3). The deterministic values of the heat capacity (dark line) are compared with the partial infinite swapping PINS65 simulations (white squares). The PINS65 results are created by averaging 22 independent samples each consisting of 10^5 moves, after a "warm up run" of 2.3×10^6 moves. These are necessary to reach convergence in the range of k_BT between 0.02 and 0.05 hartree, the left side of the heat capacity peak. The trajectories for IMD are propagated by the second order variational integrator of Sec. III with $\Delta t = 0.01$ a.u. and $k_{\max} = 100$. The values of $k_{\max}$ considered in Figs. 6 and 7 are larger than needed for the present applications. However, testing with larger $k_{\max}$ gives us assurance that our approach will be applicable to more complex problems than those considered here. From the results in Fig. 7 we can conclude that the PINS65 reaches the desired accuracy with a reasonably short simulation.

V. CONCLUSIONS

We succeed in accomplishing the two objectives set forth for the present work. We show that the INS and PINS methods can be extended to curved spaces, and we perform several numerical tests with INS and PINS on the family of decoupled double wells potentials in flat and curved spaces which have previously proved to be challenging testing grounds. In spite of the comparison made in Fig. 2, our objective is not to compare INS or PINS with parallel tempering or replica exchange. Albeit we made some effort in defining efficiency for these classes of algorithms to better understand their general performance, in our opinion one should not attempt to simulate rare events with the PINS method without trying parallel tempering first. This is, at least, because the latter is less demanding to implement. However, the data in Figs. 2 and 7, coupled with the knowledge gained in our previous work with the $(\mathrm{DDW})_n$ system,[25] provides evidence that for those cases where PT or RE has difficulty converging, PINS may improve the situation significantly.

There are several classes of problem where either the geometry is too complex, or the set of classical equations is too stiff for dynamic simulations to be efficient sampling methods. For these, parallel tempering is the only efficient sampling strategy. One example of a class of problems relevant to molecular physics where the geometry is too complex, is the class of systems that make use of Ramachandra space.[86] The number of holonomic constraints needed to simulate a polymer chain, for example, is so numerous that the Ramachandra space is best constructed from parameters of continuous Lie groups. A similar comment can be made with large aggregates of small rigid rotors requiring ellipsoids of inertia, and systems where a combination of the two must be considered. A second example of a possibly stiff problem arises when Path Integral simulations are carried out at cold temperature. The Matzubara frequencies grow linearly with the Trotter number.[84] The main contribution we make in this article is the development of an efficient integrator for spaces generally constructed from the parameters of continuous Lie groups and implement these into RE, INS, or PINS strategies to simulate molecular aggregates with curved spaces.

The results in Fig. 2 are particularly compelling and push us to develop the methods we propose in Sec. III. For $n = 30$, there are on the order of 1×10^9 minima in the potential energy surface of Eq. (3), and 1.6×10^{10} first order saddles that connect them. These relatively large numbers create the sampling challenges that can be appreciated in Fig. 2. The PINS54 simulation converges with only 10×10^6 moves, and 10^8 potential and gradient points per temperature. The graph in Fig. 2 is the best case scenario we encounter in characterizing the thermodynamics of the multidimensional double well system. For the $n = 40$ system, for example, PINS requires a significantly longer random walk before reaching the asymptotic distributions and the growth in effort with system size does not appear to be linear.

We learn that to extend the INS and PINS methods to curved spaces, we need to develop an integration method that conserves the energy reasonably well. Our method makes no use of extended Lagrangian, propagates directly in the coordinate set used to map the manifold, has second order convergence properties in regard to energy conservation, and is derived by treating the least action principle variationally. Despite the new integrator, the numerical gain over parallel tempering is less substantial for the curved spaces when compared to the flat ones. For the numerical tests in $\mathbb{R}^{30}$ we use a fourth order algorithm; therefore one may speculate that an integrator in curved spaces with higher order convergence could further improve the INS and PINS performance measured in the present work, and this will be the subject of our investigations in the near future. Additional work beyond the refinement of integrators in curved spaces will be necessary as well. Nevertheless, the development of the PINS method in curved spaces is an important step toward achieving the larger goals formulated in the introduction of this article.

ACKNOWLEDGMENTS

The authors thank Dr. D. L. Freeman and Dr. J. D. Doll for several informative conversations regarding infinite

swapping. We would also like to express our gratitude to the anonymous reviewers for several suggestions that have improved the presentation in the manuscript. E.C. acknowledges the donors of the Petroleum Research Fund, administered by the ACS (Grant No. 48146-B6).

[1]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. M. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[2]M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Claredon Press, Oxford, 1987).
[3]D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Academic Press, New York, 1996).
[4]M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley, 2008).
[5]I. Andricioaei, J. E. Straub, and A. F. Voter, J. Chem. Phys. **114**, 6994 (2001).
[6]P. Nigra, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **122**, 114113 (2005).
[7]V. Sharapov, D. Meluzzi, and V. Mandelshtam, Phys. Rev. Lett. **98**, 105701 (2007).
[8]I. R. McDonald and K. Singer, Discuss. Faraday Soc. **43**, 40 (1967).
[9]C. Bichara, J. Gaspard, and J. C. Mathieu, Phys. Lett. A **119**, 462 (1987).
[10]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[11]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[12]P. Labastie and R. L. Whetten, Phys. Rev. Lett. **65**, 1567 (1990).
[13]D. D. Frantz, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **93**, 2769 (1990).
[14]G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974).
[15]I. Andricioaei and J. E. Straub, J. Chem. Phys. **107**, 9117 (1997).
[16]Y. Pak and M. S. Wang, J. Chem. Phys. **111**, 4359 (1999).
[17]R. Zhou and B. Berne, J. Chem. Phys. **107**, 9185 (1997).
[18]M. Falcioni and M. W. Deem, J. Chem. Phys. **110**, 1754 (1999).
[19]J. P. Neirotti, F. Calvo, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 10340 (2000).
[20]F. Calvo, J. P. Neirotti, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 10350 (2000).
[21]J. P. Neirotti, D. L. Freeman, and J. D. Doll, Phys. Rev. E **62**, 7445 (2000).
[22]C. Predescu, M. Predescu, and C. V. Ciobanu, J. Chem. Phys. **120**, 4119 (2004).
[23]D. Sabo, M. Meuwly, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **128**, 174109 (2008).
[24]N. Plattner, J. D. Doll, P. Dupuis, H. Wang, Y. Liu, and J. E. Gubernatis, J. Chem. Phys. **135**, 134111 (2011).
[25]K. Roberts, R. Sebsebie, and E. Curotto, J. Chem. Phys. **136**, 074104 (2012).
[26]E. Curotto and M. Mella, J. Chem. Phys. **133**, 214301 (2010).
[27]C. Lubombo, E. Curotto, P. E. Janeiro Barral, and M. Mella, J. Chem. Phys. **131**, 034312 (2009).
[28]A. S. Case, C. G. Heid, C. M. Western, and F. Fleming Crim, J. Chem. Phys. **136**, 124310 (2012).
[29]A. S. Case, C. G. Heid, S. H. Kable, and F. Fleming Crim, J. Chem. Phys. **135**, 084312 (2011).
[30]J. Gelman-Constantin, M. A. Carignano, I. Szleifer, E. J. Marceca, and H. R. Corti, J. Chem. Phys. **133**, 024506 (2010).
[31]L. Yu and Z.-Z. Yang, J. Chem. Phys. **132**, 174109 (2010).
[32]T. S. Almeida and B. J. Costa Cabral, J. Chem. Phys. **132**, 094307 (2010).
[33]E. Asare, A-R. Musah, E. Curotto, David L. Freeman, and J. D. Doll, J. Chem. Phys. **131**, 184508 (2009).
[34]M. Mella, J. Chem. Phys. **130**, 084108 (2009).
[35]J. Douady, F. Calvo, and F. Spiegelman, J. Chem. Phys. **129**, 154305 (2008).
[36]E. Curotto, David L. Freeman, and J. D. Doll, J. Chem. Phys. **128**, 204107 (2008).
[37]T. Luan, E. Curotto, and Massimo Mella, J. Chem. Phys. **128**, 164102 (2008).
[38]M. W. Aviles, M. L. McCandless, and E. Curotto, J. Chem. Phys. **128**, 124517 (2008).
[39]P. A. Frantsuzov and V. A. Mandelshtam, J. Chem. Phys. **128**, 094304 (2008).
[40]S. F. Langley, E. Curotto, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **126**, 084506 (2007).
[41]J. Hernandez-Rojas, B. S. Gonzalez, T. James, and D. J. Wales, J. Chem. Phys. **125**, 224302 (2006).
[42]J. Deckman and V. A. Mandelshtam, Phys. Rev. E **79**, 022101 (2009).
[43]J. Deckman, P. A. Frantsuzov, and V. A. Mandelshtam, Phys. Rev. E **77**, 052102 (2008).
[44]V. A. Mandelshtam and P. A. Frantsuzov, J. Chem. Phys. **124**, 204511 (2006).
[45]P. A. Frantsuzov, D. Meluzzi, and V. Mandelshtam, Phys. Rev. Lett. **96**, 113401 (2006).
[46]P. A. Frantsuzov and V. A. Mandelshtam, Phys. Rev. E **72**, 037102 (2005).
[47]C. Predescu, P. A. Frantsuzov, and V. A. Mandelshtam, J. Chem. Phys. **122**, 154305 (2005).
[48]P. A. Frantsuzov and V. A. Mandelshtam, J. Chem. Phys. **121**, 9247 (2004).
[49]D. Sabo, C. Predescu, J. D. Doll, and D. L. Freeman, J. Chem. Phys. **121**, 856 (2004).
[50]D. D. Frantz, J. Chem. Phys. **115**, 6136 (2001).
[51]J. P. Neirotti, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 3990 (2000).
[52]D. D. Frantz, J. Chem. Phys. **107**, 1992 (1997).
[53]E. Curotto, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **109**, 1643 (1998).
[54]D. D. Frantz, J. Chem. Phys. **105**, 10030 (1996).
[55]G. E. López, J. Chem. Phys. **104**, 6650 (1996).
[56]C. Chakravarty, J. Chem. Phys. **102**, 956 (1995).
[57]D. D. Frantz, J. Chem. Phys. **102**, 3747 (1995).
[58]D. D. Frantz, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **97**, 5713 (1992).
[59]J. E. Adams and R. M. Stratt, J. Chem. Phys. **93**, 1358 (1990).
[60]D. J. Wales and R. S. Berry, J. Chem. Phys. **92**, 4283 (1990).
[61]T. L. Beck, J. D. Doll, and D. L. Freeman, J. Chem. Phys. **90**, 5651 (1989).
[62]M. Y. Hahn and R. L. Whetten, Phys. Rev. Lett. **61**, 1190 (1988).
[63]T. L. Beck, D. M. Leitner, and R. S. Berry, J. Chem. Phys. **89**, 1681 (1988).
[64]E. Blaisten-Barojas, I. L. Garzón, and M. Avalos-Borja, Phys. Rev. B **36**, 8447 (1987).
[65]J. Jellinek, T. Beck, and R. S. Berry, J. Chem. Phys. **84**, 2783 (1986).
[66]D. L. Freeman and J. D. Doll, J. Chem. Phys. **82**, 462 (1985).
[67]N. G. Garcia and J. M. S. Torroja, Phys. Rev. Lett. **47**, 186 (1981).
[68]M. R. Hoare and P. Pal, Adv. Phys. **20**, 161 (1971).
[69]P. J. Rossky, J. D. Doll, and H. L. Friedman, J. Chem. Phys. **69**, 4628 (1978).
[70]M. Rao and B. J. Berne, J. Chem. Phys. **71**, 129 (1979).
[71]P. V. Kumar, J. S. Raut, S. J. Warakomski, and K. A. Fichthorn, J. Chem. Phys. **105**, 686 (1996).
[72]H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[73]T. A. Andrea, W. C. Swope, and H. C. Andersen, J. Chem. Phys. **79**, 4576 (1983).
[74]S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[75]W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982).
[76]Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[77]R. D. Mountain and D. Thirumalai, J. Phys. Chem. **93**, 6975 (1989).
[78]Raw data is available upon request from the author.
[79]J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[80]H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).
[81]W. D. Curtis and F. R. Miller, *Differential Manifolds and Theoretical Physics* (Academic Press, 1985).
[82]S. Nair, Physica D **241**, 315 (2012).
[83]H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* (World Scientific, Singapore, 1990).
[84]S. Wolf and E. Curotto, J. Chem. Phys. **137**, 014109 (2012).
[85]R. Gillilan and K. Wilson, J. Chem. Phys. **97**, 1757 (1992).
[86]A. Patriciu, G. S. Chirikjian, and R. V. Pappu, J. Chem. Phys. **121**, 12708 (2004).