

On the efficient sampling of pathways in the transition path ensemble

Thijs J. H. Vlugt[†] and Berend Smit^{*}

Department of Chemical Engineering, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands. E-mail: smit@its.chem.uva.nl

Received 8th December 2000, Accepted 12th January 2001
Published on the Web 25th January 2001

Transition path sampling is a numerical technique to compute transition rates between two stable states without having to assume *a priori* information about a dividing surface. However, when there are many different pathways to go from one stable state to another, for example, when there are many saddle points in the free energy surface, it can be difficult to sample all possible pathways within the timescale of a single simulation. In this work, we demonstrate the use of parallel tempering to overcome this problem.

1 Introduction

Many processes in nature, like chemical reactions for example, are rare events. Many computational techniques are available for studying this class of processes, of which transition state theory¹ is best known. This technique requires *a priori* information about the so-called transition state which is usually defined as a saddle point in the free energy surface. There are many numerical techniques available to locate these saddle points (see, for example, ref.2 and references therein). However, these techniques are often limited to systems with only a few saddle points.

Recently, Dellago, Bolhuis and Chandler^{3–5} have introduced transition path sampling (TPS). This technique is able to compute rate constants without having to assume a certain transition state. As a Monte Carlo (MC) procedure is applied to sample deterministic trajectories between two stable sites, the system itself will explore all possible pathways, *i.e.* pass all possible saddle points of the free energy surface. This technique has been used to compute rate constants in complex chemical systems, *e.g.*, ion dissociation in water,⁶ isomerization of biomolecules,⁷ and diffusion of hydrocarbons in zeolites.⁸

Although TPS requires that we sample all possible transition paths, we still have to make sure that our MC algorithm is able to do this within a finite amount of MC steps; *i.e.* if there are different transition paths that are not connected in path space problems can be expected. This is for example the case when there is an energy barrier between two saddle points which is much larger than $k_B T$, see Fig. 1. When a path is passing one saddle point, it will be very unlikely that this path will evolve to a path that is passing the other saddle point.

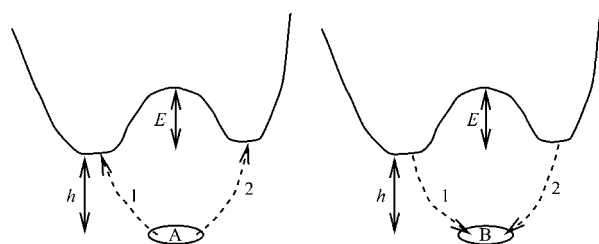


Fig. 1 Schematic representation of two different transition paths from site A to site B (left: front, right: back). A and B are two stable sites, separated by an energy barrier of size h . There are two dynamical pathways for the system to go from A to B (1 and 2). Due to the large energy barrier E between these paths, it will be very unlikely that path 1 will evolve to path 2 in a single MC simulation. Note that the energy barrier between two paths (E) is different from the energy barrier along the path (here, $h=6$).

In this paper, we will demonstrate how to solve this problem using parallel tempering (PT).⁹ With this technique, several systems with a slightly different partition function are sampled in the same simulation. Besides conventional MC trial moves, attempts are made to exchange configurations between different systems. The idea is that transitions, which without exchange moves cannot occur in a particular system, do occur because the transition is allowed in another system. This technique has been found very useful in many applications, *e.g.*, the determination of zeolite structures,¹⁰ computation of phase equilibria,^{11,12} and the simulation of polymer melts.¹³

The outline of this paper is as follows. In section 2, we will briefly discuss the TPS and PT techniques. In section 3, we will discuss a simple model system for which we will demonstrate the usefulness of parallel tempering in TPS simulations in section 4.

2 Simulation method

In this section we will give a short introduction to transition path sampling. We will discuss how to apply parallel tempering to transition path sampling.

2.1 Transition path sampling

The calculation of the transition rate using the transition path ensemble (TPE) technique consists of several steps (details of this technique can be found in ref. 4 and 5):

1. Define the function $h_R(x)=1$ when $x \in R$ and $h_R(x)=0$ otherwise, where R is either region A or B. This defines the two stable sites. The transition from A to B should be rare.

2. Compute the probability $\langle h_B(t) \rangle$ that a path of length T starting in A ends in B after a time t provided that it has been in B at least once during the time interval $[0, T]$:

$$\begin{aligned} \langle h_B(t) \rangle &= \frac{\int dx_0 \exp[-\beta \mathcal{H}(x_0)] h_A(x_0) H_B(x_0, T) h_B(x_t)}{\int dx_0 \exp[-\beta \mathcal{H}(x_0)] h_A(x_0) H_B(x_0, T)} \\ &= \frac{\int dx_0 F(x_0, T) h_B(x_t)}{\int dx_0 F(x_0, T)} \end{aligned} \quad (1)$$

in which $t \in [0, T]$ and

$$H_B(x_0, T) = \max_{0 \leq t \leq T} h_B(x_t) \quad (2)$$

$$F(x_0, T) = \exp[-\beta \mathcal{H}(x_0)] h_A(x_0) H_B(x_0, T) \quad (3)$$

Here, we have assumed the following:

(a) All paths are deterministic, *i.e.* the position of the end of the path (x_t) follows directly from the initial condition x_0 by integrating the equations of motion, *i.e.* $x_t = x_t(x_0)$. We will use

molecular dynamics (MD) in the microcanonical (NVE) ensemble for this.

(b) The initial conditions x_0 are taken from a Boltzmann distribution, *i.e.* the total energy of a path (which is constant along a path because the path itself is obtained by a MD simulation in the NVE ensemble) is sampled in the canonical ensemble.

3. Compute the probability $C(t)$ that a path that starts in A ends in B after time t . For this, we can write

$$C(t) = \frac{\int dx_0 \exp[-\beta \mathcal{H}(x_0)] h_A(x_0) h_B(x_t)}{\int dx_0 \exp[-\beta \mathcal{H}(x_0)] h_A(x_0)} = \int_{\lambda_{\min}}^{\lambda_{\max}} d\lambda P(\lambda, t) \quad (4)$$

in which

$$P(\lambda, t) = \frac{\int dx_0 \exp[-\beta \mathcal{H}(x_0)] h_A(x_0) \delta[\lambda - \lambda(x_t)]}{\int dx_0 \exp[-\beta \mathcal{H}(x_0)] h_A(x_0)} \quad (5)$$

Here, we have defined region B using an order parameter λ in such a way that region B is between λ_{\min} and λ_{\max} . $P(\lambda, t)$ can be interpreted as the probability for the system to be in a state with a certain λ after time t given that the system is in A at time 0. Because $P(\lambda, t)$ is quite small in B (*i.e.* transitions from A to B are rare), it is advantageous to use umbrella sampling¹⁴⁻¹⁶ to compute $P(\lambda, t)$. By defining overlapping regions B_i by

$$x_t \in B_i \quad \text{if} \quad \lambda_{\min}(i) \leq \lambda(x_t) \leq \lambda_{\max}(i) \quad (6)$$

in such a way that the union of B_i equals the total phase space, one is able to calculate

$$\begin{aligned} P(\lambda, t, i) &= \frac{\int dx_0 \exp[-\beta \mathcal{H}(x_0)] h_A(x_0) h_{B_i}(x_t) \delta[\lambda - \lambda(x_t)]}{\int dx_0 \exp[-\beta \mathcal{H}(x_0)] h_A(x_0) h_{B_i}(x_t)} \\ &= \frac{\int dx_0 f(x_0, t, i) \delta[\lambda - \lambda(x_t)]}{\int dx_0 f(x_0, t, i)} \end{aligned} \quad (7)$$

in which

$$f(x_0, t, i) = \exp[-\beta \mathcal{H}(x_0)] h_A(x_0) h_{B_i}(x_t) \quad (8)$$

$f(x_0, t, i)$ is the ensemble of all paths starting in A and ending in B_i at time t . Because $P(\lambda, t, i) \propto P(\lambda, t, j)$, one is able to construct $P(\lambda, t)$ by matching the histograms and normalizing.

4. Finally, the transition rate $k_{A \rightarrow B}$ (number of events per unit of time) is calculated using

$$k_{A \rightarrow B} = \frac{dC(t)}{dt} = \frac{C(t)}{\langle h_B(t) \rangle_{F(x_0, T)}} \times \frac{d[\langle h_B(t) \rangle_{F(x_0, T)}]}{dt} \quad (9)$$

The quantities $\langle h_B(t) \rangle$ [eqn. (1)] and $P(\lambda, i)$ [eqn. (7)] can be interpreted as ensemble averages over distributions $F(x_0, T)$ and $f(x_0, t, i)$, respectively. As these ensemble averages are averages over paths starting in A (represented by the initial condition x_0), the resulting ensemble is called the transition path ensemble. Therefore, one can use a conventional MC procedure to sample from these distributions. The trial moves for sampling from these distributions are described in detail in ref.5 and briefly mentioned in section 3.

2.2 Parallel tempering

Parallel tempering^{9-11,17} is a very useful MC technique for systems that suffer from many local minima in which the system might get stuck. In this technique, N independent systems are simulated simultaneously. The total partition function of this system (Q) equals

$$Q = \prod_{i=1}^N Q_i \quad (10)$$

in which

$$Q_i = \int dx_i f_i(x_i, \beta_i) \quad (11)$$

For the canonical ensemble, the function $f(x_i, \beta_i)$ equals

$$f_i(x_i, \beta_i) = \exp[-\beta_i \mathcal{H}_i(x_i)] \quad (12)$$

in which $\beta_i = 1/(k_B T_i)$. For each of these systems, individual trial moves are performed. After a randomly selected number of trial moves, an attempt is made to exchange configurations. Two systems (i and j , $|i-j|=1$) are selected at random, the systems are exchanged by choosing $x_i(n) = x_j(o)$ and $x_j(n) = x_i(o)$ in which we have used the symbols 'n' and 'o' for the new and old configuration, respectively. The ratio of acceptance probabilities equals

$$\frac{\text{acc}(o \rightarrow n)}{\text{acc}(n \rightarrow o)} = \frac{\prod_{k=1}^N f_i(x_k(n), \beta_k)}{\prod_{k=1}^N f_i(x_k(o), \beta_k)} = \frac{f_i(x_j(o), \beta_i) f_j(x_i(o), \beta_j)}{f_i(x_i(o), \beta_i) f_j(x_j(o), \beta_j)} \quad (13)$$

Such trial moves will be accepted when there is enough overlap between $f_i(x, \beta_i)$ and $f_j(x, \beta_j)$. Usually, the temperature is used as a control parameter to distinguish between different systems i and j , which implies $\mathcal{H}_i(x) = \mathcal{H}_j(x)$ for all x . This is a convenient choice, because one is usually interested in computing physical properties at different temperatures. However, this does not necessarily have to be the case. An example of this is the use of N potentials that slightly differ in softness¹³ in such a way that the limiting potentials are the ideal gas and the full excluded volume potential. However, in TPS simulations this approach is not allowed because it would destroy the determinism of the paths. Therefore, we will use in the remaining part of this article $\mathcal{H}_i(x) = \mathcal{H}_j(x) = \mathcal{H}(x)$.

We will discuss two ways of using parallel tempering in the TPE:

1. Exchanges between different temperatures. When paths are sampled in the canonical ensemble, one can simulate different systems with different β and apply exchange trial moves between different systems. This means that a transition between two classes of pathways that, without parallel tempering, rarely occurs in system i but quite frequently in system j (for example, because $T_j \gg T_i$), can also occur in system i . As the two paths that are exchanged both have the same constraints for the start and the end of the path, the acceptance/rejection rule [eqn.(13)] reduces to

$$\frac{\text{acc}(o \rightarrow n)}{\text{acc}(n \rightarrow o)} = \exp[(\beta_i - \beta_j) \times (\mathcal{H}(x_i(o)) - \mathcal{H}(x_j(o)))] \quad (14)$$

2. Exchanges between different regions B_i (umbrella sampling). To compute $P(\lambda, t)$, it has been shown in the previous section that it is advantageous to define overlapping regions B_i in such a way that $\cup B_i$ equals the whole phase space. When all regions are simulated simultaneously, one is able to perform trial moves that exchange paths between two overlapping regions B_i and B_j ($i \neq j$). Such a trial move is only accepted when the end points of both paths are in the overlapping region of B_i and B_j (*i.e.* $h_{B_i}(x_t) = h_{B_j}(x_t) = 1$). A similar application of parallel tempering in the context of umbrella sampling has been used by Auer and Frenkel to compute the free energy barrier for homogeneous crystal nucleation of hard-sphere colloids.¹⁸

In the remaining part of this article, we will demonstrate the efficiency of both schemes, as well as the combination of both.

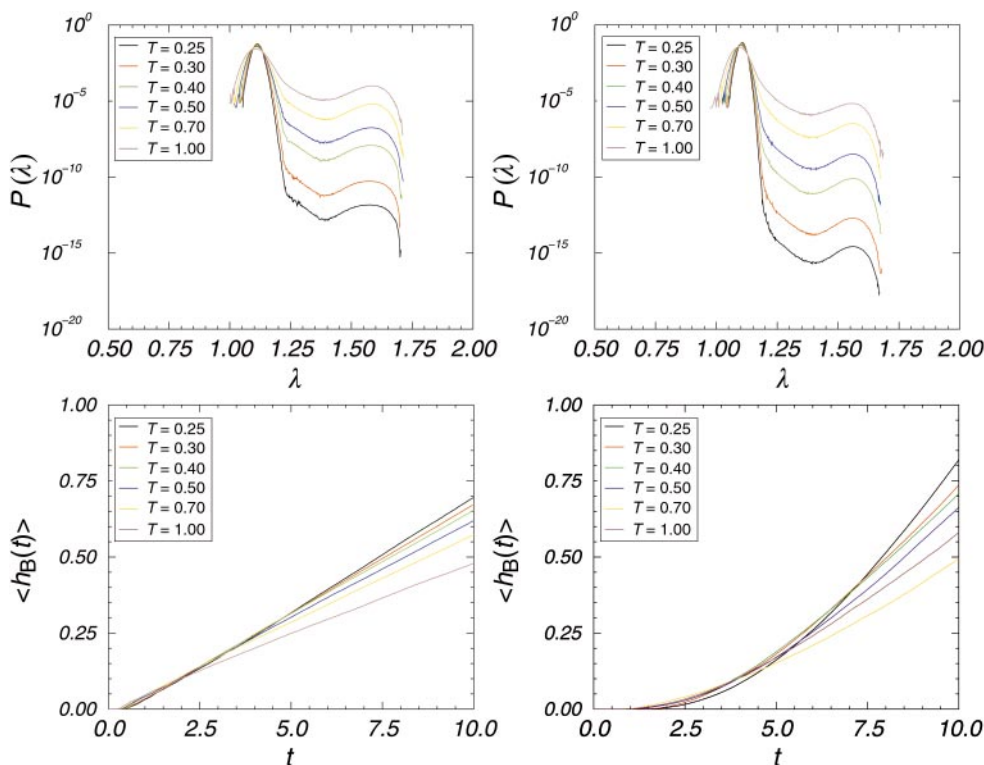


Fig. 2 Top: the functions $P(\lambda, t)$ for $t=5$. Bottom: the function $\langle h_B(t) \rangle$ for $T=10$. Left: $N_s=2$. Right: $N_s=4$.

3 Model system

To illustrate this method, we have used a two-dimensional system consisting of 15 WCA (Weeks–Chandler–Andersen) particles:^{5,19}

$$u_{\text{WCA}}(r) = \begin{cases} 1 + 4[r^{-12} - r^{-6}] & r \leq r_{\text{WCA}} \\ 0 & r > r_{\text{WCA}} \end{cases} \quad (15)$$

in which r is the distance between two particles and $r_{\text{WCA}} \equiv 2^{1/6}$. However, N_s particle pairs (1–2, 2–3, ..., $N_s - (N_s + 1)$) interact via a double well potential to form a chain of length $N_s + 1$:

$$u_{\text{dw}}(r_i) = h \left[1 - \frac{(r_i - w - r_{\text{WCA}})^2}{w^2} \right]^2 \quad (16)$$

in which r_i is the distance between particles i and $i+1$. This potential has two minima for $r_i = r_{\text{WCA}}$ and $r_i = r_{\text{WCA}} + 2w$. We have used $w=0.25$ and $h=6$. The system is confined in a circle of radius 3.0 for which a WCA potential at the boundary is used, this corresponds to a reduced density of 0.6. The equations of motion are integrated using the (time-reversible) velocity-Verlet algorithm with a time-step of 0.001. We have chosen the regions A and B such that $x \in A$ when $(r_1, r_2, \dots, r_{N_s}) < 1.30$ and $x \in B$ when $\min(r_1, r_2, \dots, r_{N_s}) > 1.45$. In other words, in state A all bonds are in the compact state while in state B all bonds are in the extended state. As the barrier height is equal for all bonds, we may expect that we can express the transition rate from A to B by a single rate constant [eqn.(9)]. As different bonds can stretch at a different time, there are $(N_s - 1)!$ possible pathways for which this chain transforms from a compact state to form an extended state. A natural order parameter for studying these different paths would be the identity of the bond that it stretched first. We will call this order parameter X . Because of the symmetry of the system the average value of this order parameter equals

$$\langle X \rangle = \frac{N_s + 1}{2} \quad (17)$$

For the calculation of $P(\lambda)$, we are not guaranteed that the bonds are in the extended state (depending on the region B_i). Therefore, we have used $\lambda = (r_1, r_2, \dots, r_{N_s})$ as an order parameter. The length of the transition paths is $T=10.0$, the function $P(\lambda)$ has been calculated at half this time.

We have performed our simulations at 6 temperatures simultaneously, $T_1=0.25$, $T_2=0.30$, $T_3=0.40$, $T_4=0.50$, $T_5=0.70$, and $T_6=1.00$. Additionally, for the calculation of $P(\lambda)$ we have defined 12 overlapping regions: $0.00 < \lambda_1 < 1.25$, $1.14 < \lambda_2 < 1.25$, $1.15 < \lambda_3 < 1.25$, $1.16 < \lambda_4 < 1.25$, $1.17 < \lambda_1 < 1.25$, $1.18 < \lambda_2 < 1.25$, $1.19 < \lambda_3 < 1.30$, $1.20 < \lambda_4 < 1.30$, $1.21 < \lambda_4 < 1.30$, $1.23 < \lambda_5 < 1.30$, $1.25 < \lambda_6 < 1.35$, and $1.30 < \lambda_7 < \infty$. This means that for the calculation of $P(\lambda)$, $6 \times 12 = 72$ systems are simulated simultaneously.

In our simulations, there are three types of (randomly selected) trial moves:

1. Shifting $[(70-x)\%]$. In this move, the initial conditions of a (randomly selected) path are translated by an amount Δt . Due to the energy conservation of a trajectory the acceptance rule for this trial move equals for the calculation of $\langle h_B(t) \rangle$

$$\text{acc}(o \rightarrow n) = h_A(x_0(n))H_B(x_0(n), T) \quad (18)$$

and for the calculation of $P(\lambda, i)$

$$\text{acc}(o \rightarrow n) = h_A(x_0(n))h_B(x_i(n)) \quad (19)$$

2. Shooting (30%). A randomly selected particle of a randomly selected slice of a randomly selected system is given a random displacement; the maximum displacement is adjusted such that 20% of the trial moves are accepted.

3. Exchange of configurations ($x\%$). Two overlapping systems are selected at random. An attempt is made to exchange these systems. For the calculation of $\langle h_B(t) \rangle$ we have used exchanges between different temperatures only. For the calculation of $P(\lambda)$, we have chosen at random with equal probability to exchange configurations between different temperatures (β_i, β_j) or between overlapping regions (B_i, B_j). Note that the last trial move is also possible when the ensemble in which the paths are sampled is the micro-canonical

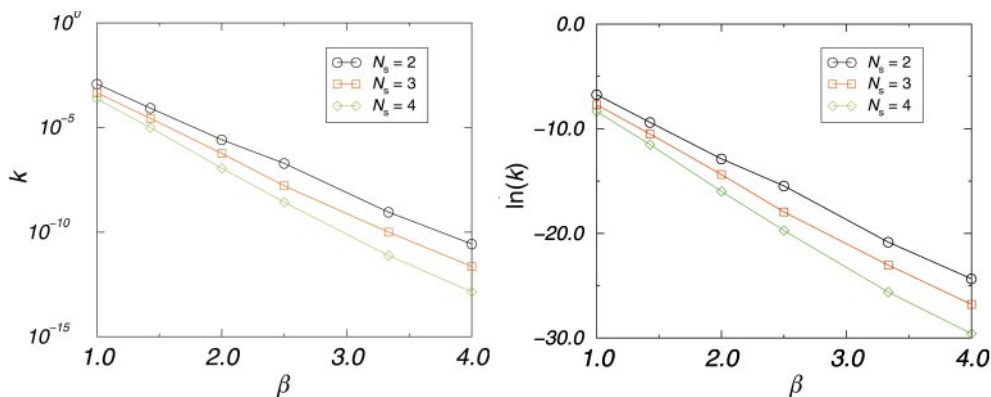


Fig. 3 Transaction rate k (events per unit of time) as a function of β . The slope of the right curve equals the activation energy, which is 5.9 for $N_s=2$, 6.3 for $N_s=3$, and 7.0 for $N_s=4$. These numbers are almost equal to the size of the barrier between the compact and the extended state of the double well potential ($h=6$).

ensemble, *i.e.*

$$f(x_0, t, i) = \delta(\mathcal{H} - \mathcal{H}_0) h_A(x_0) h_{B_i}(x_t) \quad (20)$$

in which \mathcal{H}_0 is the constant total energy of a path.

The total simulation consisted of 2×10^6 cycles; in every cycle the number of trial moves equals the number of slices multiplied with the number of temperatures. We found that our particular choice of Hamiltonians leads to an acceptance probability for exchange of configurations between 0.02 and 0.20.

4 Results and discussion

To compare efficiencies of different types of simulations, we have computed the autocorrelation functions of the total energy of the paths as well as the autocorrelation of the identity of the path. The autocorrelation function of quantity A equals

$$G(t) = \frac{\langle \delta A(0) \delta A(t) \rangle}{\langle \delta A^2 \rangle} \quad (21)$$

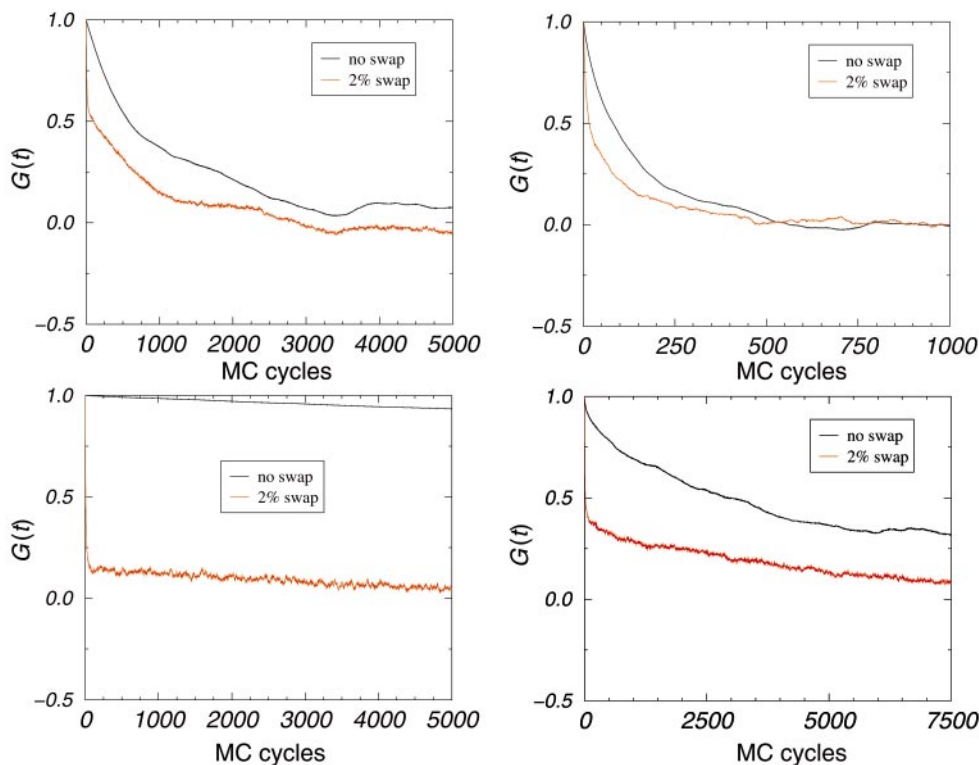


Fig. 4 Autocorrelation functions for the energy (left) and order parameter (X , right) for $N_s=2$ (top) and $N_s=4$ (bottom).

in which $\delta A = A - \langle A \rangle$ and t is the number of simulation cycles. To investigate the efficiency of a simulation, we will only discuss the autocorrelation at the lowest temperature. This is because equilibration of the system is most difficult at the lowest temperature.

4.1 Calculation of the transition rate

In Fig. 2, we have plotted the functions $P(\lambda)$ and $\langle h_B(t) \rangle$ for $N_s=2$ and $N_s=4$. Clearly, one is able to see the maximum in $P(\lambda)$ around $\lambda = r_{WCA} \approx 1.12$, which indicates that most paths which start in A will stay in A on short timescales and therefore the transition from A to B is a rare event. A similar maximum is observed for region B when $\lambda = r_{WCA} + 2w \approx 1.62$. The function $P(\lambda)$ decreases many orders of magnitude for large λ when the temperature is lowered. Therefore, we need to use umbrella sampling to compute this function. For $t > 5$, the function $\langle h_B(t) \rangle$ clearly approaches a straight line for all systems. This is a requirement to compute the hopping rate [see eqn.(9)].

In Fig. 3, we have plotted the transition rate k as a function of

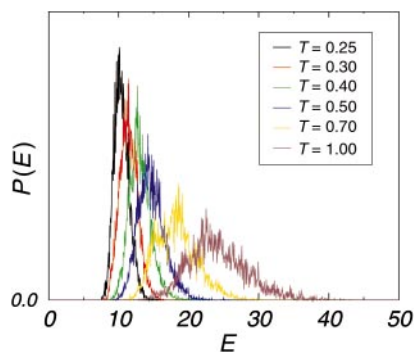


Fig. 5 Distribution of the total energy for $N_s=2$ for various temperatures. As these distributions have a considerable overlap, exchanges of configurations (parallel tempering) can be applied successfully.

β . Apparently, $\ln(k)$ is proportional to β (Arrhenius' law). The slope of these graphs equals the activation energy, which is roughly equal to the height of the barrier between the compact and the extended state of the double well potential (here, $h=6$). Note that the hopping rate is lower for larger values of N_s .

4.2 Calculation of $\langle h_B(t) \rangle$

In Fig. 4, we have plotted the autocorrelation functions of the energy and the order parameter X for $N_s=2$ and $N_s=4$. For $N_s=2$, there is not much difference in efficiency when parallel tempering is applied (2% swap). For $N_s=4$, however, the differences are huge. When parallel tempering is applied for this system, the autocorrelation functions for the energy and order parameter X decay very fast initially, after which a somewhat slower relaxation is observed. Note that when no parallel tempering is applied, it is almost impossible to sample the system efficiently for $N_s=4$.

In Fig. 5, we have plotted the distribution of the total energy of the path for all temperatures. This distribution broadens significantly when the temperature increases. To have a successful acceptance rate of exchange moves, it is essential that the energy distributions have a considerable overlap.

4.3 Calculation of $P(\lambda)$

In Fig. 6 and 7, we have plotted the autocorrelation functions for the energy and order parameter λ for the first, middle, and last slice for both $N_s=2$ as well as for $N_s=4$. The black line indicates the autocorrelation function without the use of parallel tempering. The other lines indicate the use of parallel

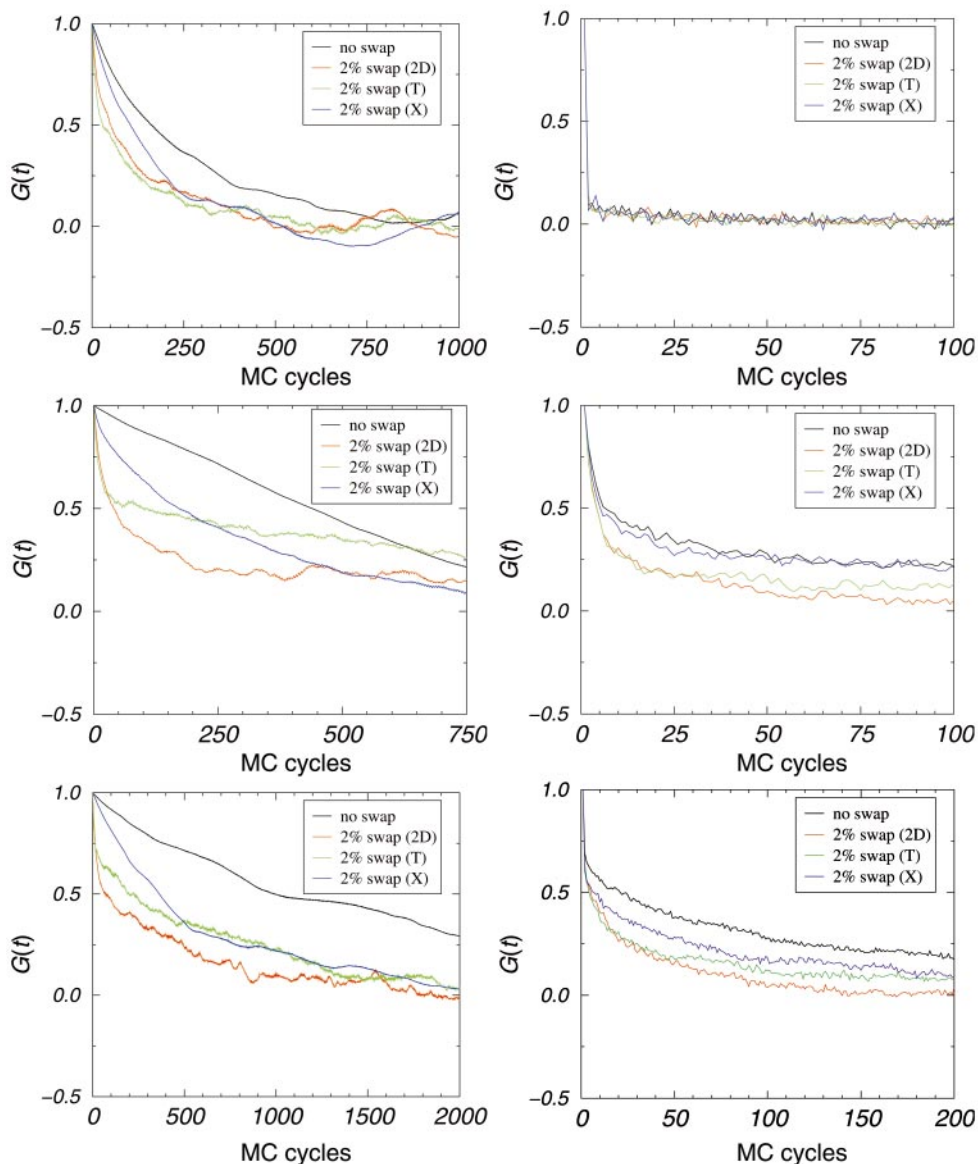


Fig. 6 Autocorrelation functions for the energy (left) and order parameter (λ , right) for $N_s=2$. Top: first slice ($i=1$), middle: middle slice ($i=6$), bottom: last slice ($i=12$).

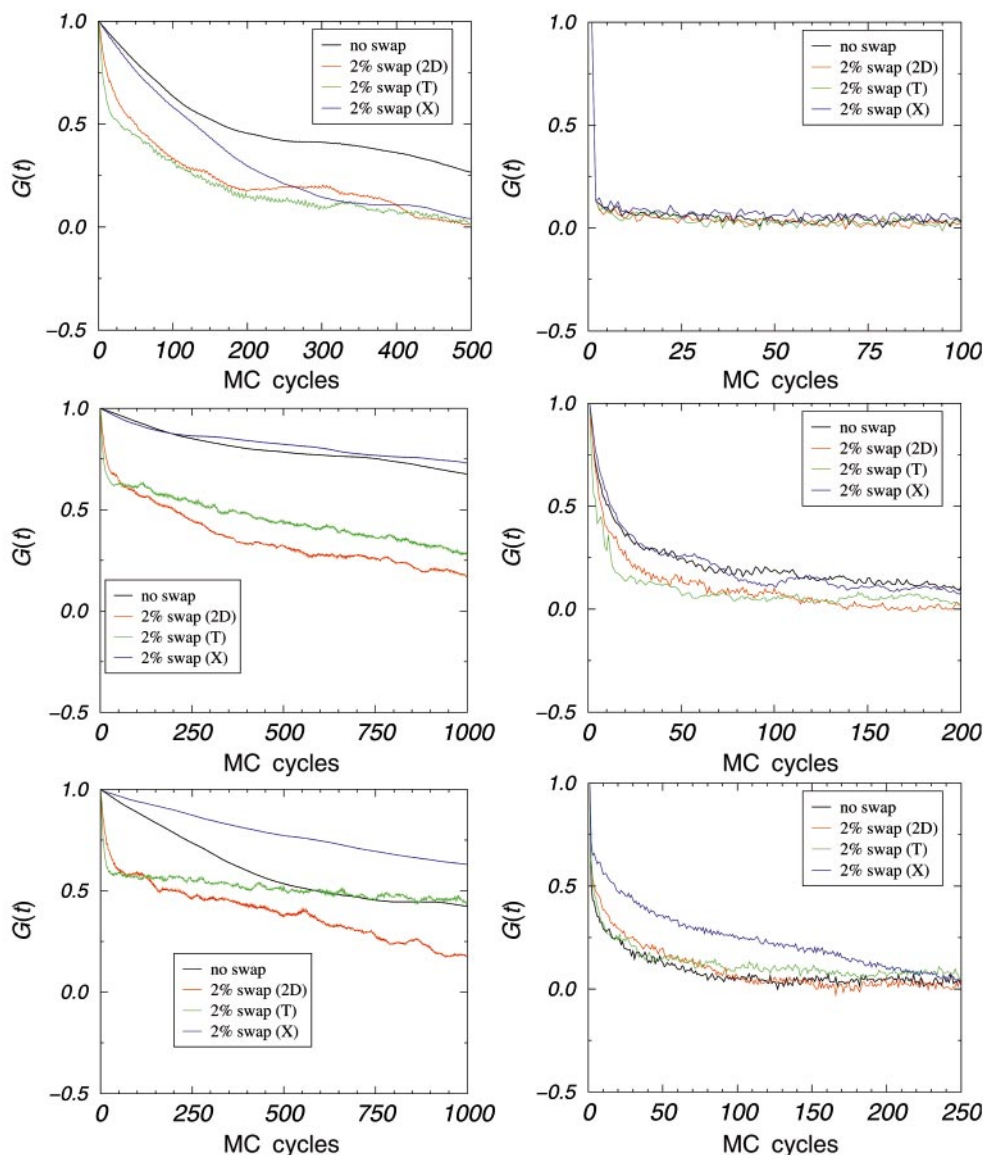


Fig. 7 Autocorrelation functions for the energy (left) and order parameter (λ , right) for $N_s=4$. Top: first slice ($i=1$), middle: middle slice ($i=6$), bottom: last slice ($i=12$).

tempering in two directions (red), only for different temperatures (green), and only for different regions B_i (blue).

For all simulations, the correlation time of the total energy is

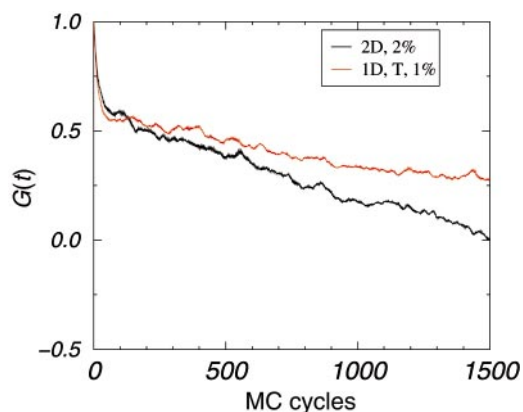


Fig. 8 Autocorrelation functions for the energy for slice 13 ($N_s=4$). Comparison between 2% exchange moves (50% temperature, 50% slices, red) and 1% exchange moves in temperature (black). This figure demonstrates that applying parallel tempering in more dimensions is most efficient.

much larger than the correlation time of λ . The correlation time of λ is very small for the first slice both for $N_s=2$ as well as for $N_s=4$. For the first slice, parallel tempering in only one direction (temperature) is most efficient, while parallel tempering both in temperature as in slices is more efficient for all other slices. As these slices have a larger correlation time, they determine the overall correlation time of the simulation.

To demonstrate the effect of swapping between different slices, in Fig. 8 we have plotted the energy autocorrelation function for $N_s=4$ with and without exchanges between slices, while the fraction of attempted exchanges between different temperatures is constant (0.01). Again, this demonstrates the effective use of exchanges in different dimensions.

5 Conclusions

In summary, we have used parallel tempering to sample the transition path ensemble more efficiently and to overcome ergodicity problems. It was found that a combination of directions in which parallel tempering is applied is most efficient. This suggests that one should set up a simulation in which all systems of interest are simulated simultaneously, instead of the traditional approach of separate simulations for each system of interest.

Acknowledgements

Financial support was provided to TJHV by NWO-CW (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Chemische Wetenschappen). BS acknowledges support from NOW-CW through PIONIER. The authors would like to thank Cameron F. Abrams for a critical reading of the manuscript.

References

- 1 D. Chandler, *J. Chem. Phys.*, 1978, **68**, 2959.
- 2 G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 1999, **111**, 7010.
- 3 C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.*, 1998, **108**, 1964.
- 4 P. G. Bolhuis, C. Dellago and D. Chandler, *Faraday Discuss.*, 1998, **110**, 421.
- 5 C. Dellago, P. G. Bolhuis and D. Chandler, *J. Chem. Phys.*, 1999, **110**, 6617.
- 6 P. L. Geissler, C. Dellago and D. Chandler, *J. Phys. Chem. B*, 1999, **103**, 3706.
- 7 P. G. Bolhuis, C. Dellago and D. Chandler, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 5877.
- 8 T. J. H. Vlugt, C. Dellago and B. Smit, *J. Chem. Phys.*, 2000, **113**, 8791.
- 9 C. J. Geyer and E. A. Thompson, *J. Am. Stat. Assoc.*, 1995, **90**, 909.
- 10 M. Falcioni and M. W. Deem, *J. Chem. Phys.*, 1999, **110**, 1754.
- 11 Q. L. Yan and J. J. de Pablo, *J. Chem. Phys.*, 1999, **111**, 9509.
- 12 Q. L. Yan and J. J. de Pablo, *J. Chem. Phys.*, 2000, **113**, 1276.
- 13 A. Bunker and B. Dünweg, *Phys. Rev. E*, 2001, **63**, 1670.
- 14 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187.
- 15 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
- 16 D. Frenkel and B. Smit, *Understanding Molecular Simulations: from Algorithms to Applications*, Academic Press, San Diego, 1996.
- 17 M. G. Wu and M. W. Deem, *Mol. Phys.*, 1999, **97**, 559.
- 18 S. Auer and D. Frenkel, *Nature*, 2001, in the press.
- 19 J. D. Weeks, D. Chandler and H. C. Andersen, *J. Chem. Phys.*, 1971, **54**, 5237.