

A simple method for sampling partition function ratios

Andreas Schüring^a, Scott M. Auerbach^b, Siegfried Fritzsche^{a,*}

^a *University of Leipzig, Faculty of Physics and Geosciences, Institute of Theoretical Physics, Postfach 100920, 04009 Leipzig, Germany*

^b *Department of Chemistry and Department of Chemical Engineering, University of Massachusetts, Amherst, MA 01003, USA*

Received 25 July 2007; in final form 2 November 2007

Available online 7 November 2007

Abstract

A simple simulation method is proposed for calculating partition function ratios with emphasis on computing transition state theory rate constants. Potential energy histograms are calculated in one single high temperature simulation. Partition function ratios can then be computed for a wide range of temperatures from this information. We apply this high-temperature configuration-space exploration (HTCE) method to a model problem. We find that HTCE quantitatively reproduces rate constants evaluated using brute-force molecular dynamics, and using Voter's displacement vector method. HTCE is extremely straightforward to implement and it offers other system information that targeted free-energy simulations may not.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Transition state theory (TST) provides the conceptual kinetic picture underlying activated molecular processes [1], including chemical reactions, protein folding and diffusion in solids [2–4] where it can form the basis for hierarchical simulations covering large timescales [5,6].

TST is applied to systems with two stable regions A and B of low potential energy (or local free energy) while moving from A to B the system has to pass a region of high potential energy (free energy). Hence, the system will spend most of the time in state A or B, respectively and the passage is a rare event. A hypersurface dividing the configuration space into these regions A and B is called the dividing surface S. TST calculates the flux passing through the surface S. The possibility that the system after passing the surface S can still move back to state A must be taken into account by a dynamical correction factor which is to be calculated additionally to TST. State A is called the reactant region, state B the product region and the so called transition state, marked by #, is separating both states. The most

probable path from A to B is called the minimum energy path (MEP).

Despite its simple picture of thermally-activated barrier crossing, implementing TST as a computational tool for condensed phase systems remains challenging, because of the need for special sampling techniques that target the dividing surface [7]. A nice review of recent advances in implementing TST for many-body systems can be found in Ref. [7]. Many of these methods, such as hyperdynamics [8] and transition path sampling [9], were created to bypass the specification of a reaction coordinate/dividing surface, because such constructs may be difficult to visualize in many-body systems that exhibit cooperative dynamics. Even when a naive reaction coordinate can be specified, such as for a molecule executing cage-to-cage motion in a zeolite, computing the relevant ratio of partition functions can be challenging because of the need for special importance sampling. In this Letter, we report an extremely straightforward simulation approach for modeling rare events that produces TST rates over a wide temperature range.

Special sampling techniques that target the dividing surface as in [7] or the basin-constrained MD simulation [10] are not necessary. No special reaction pathways and energy thresholds are examined to find a system of rate constants that are scaled by Boltzmann weights as it is done in [10].

* Corresponding author. Fax: +49 341 2352307.

E-mail address: siegfried.fritzsche@uni-leipzig.de (S. Fritzsche).

Instead, only straightforward MD simulation without any biasing is required.

Given a reaction coordinate, TST rates are typically computed using umbrella sampling with Monte Carlo (MC) or some other sampling method [11]. An example is Voter's displacement vector Monte Carlo method [12], which works particularly well for modeling adatom diffusion on surfaces, and has been adapted to modeling diffusion in zeolites [13]. It is worthwhile to ask whether straightforward simulation, i.e., without specialized umbrella sampling, can also provide TST rates over wide temperature ranges. This question is warranted in part by the simplicity of molecular dynamics (MD) [11] and the availability of many robust MD codes [14]. Such an approach may make TST calculations accessible to a wider population of the molecular modeling community.

In what follows, we report that high-temperature MD with potential energy histogramming can yield statistically accurate TST rates over a wide temperature range, including low temperatures. The benefit of this method, which we call high-temperature configuration-space exploration (HTCE), is that from a single high-temperature simulation, enough information can be accumulated to construct partition function ratios at many temperatures, as we show below. In addition, during an HTCE run many other system properties can be monitored. Our approach stands in contrast to constrained MD approaches for computing partition function ratios [15,16], which essentially use constrained MD instead of MC for umbrella sampling. Instead, our approach can essentially be considered to be a special dynamical version of the very wide class of histogram reweighting procedures developed initially by Ferrenberg and Swendsen for MC of bulk liquids close to the critical point and proposed in [17] for different system parameters.

Our approach, which uses totally unconstrained MD at high temperatures, yields a remarkably simple approach for implementing TST computations in materials science.

The remainder of this Letter is organized as follows: in Section 2 we describe the theory behind HTCE; in Section 3 we show example simulation results; in Section 4 we discuss convergence of the HTCE method; and in Section 5 we offer concluding remarks.

2. Theory

Transition state theory (TST) is applied to systems with two stable regions A and B of low potential energy (or local free energy) while moving from A to B the system has to pass a region of high potential energy (free energy). Hence, the system will spend most of the time in state A or B respectively and the passage is a rare event. A hypersurface dividing the configuration space into these regions A and B is called the dividing surface S. TST calculates the flux passing through the surface S. The possibility that the system after passing the surface S can still move back to state A must be taken into account by a dynamical correction factor which is to be calculated additionally to TST. State

A is called the reactant region, state B the product region and the so-called transition state, marked by #, is separating both states.

According to TST, the rate constant k for the transition can be estimated from the formula:

$$k^{\text{TST}} = \frac{1}{2} \langle |v| \rangle \lim_{\delta \rightarrow 0} \frac{1}{\delta} \frac{P_{\#}(\delta)}{P_A}, \quad (1)$$

where P_A and $P_{\#}(\delta)$ are the probabilities to find the system in reactant state A, or in a small shell of thickness δ at the transition state dividing surface #, respectively. In Eq. (1), $0.5\langle |v| \rangle$ is the average speed perpendicular to the dividing surface in the direction from A to #. Using straightforward MC or MD at ambient temperatures to sample the relative probabilities of visiting regions A and # is obviously fraught with trouble because even in very long MD runs the transition region will hardly be visited at all and a satisfactory statistics cannot be obtained. However, at high temperatures the system visits the transition region much more frequently. Our basic idea is to calculate probabilities and density of states at high temperatures, then use these to construct the ratio of the probabilities P_A and $P_{\#}$ at ambient temperatures. In what follows, we designate the high temperature used for sampling as T^* (with $\beta^* = 1/k_B T^*$), while ambient temperatures are designated generally as T ($\beta = 1/k_B T$).

In the canonical ensemble the configurational part Z of the canonical partition function reads in M dimensions

$$Z_M = \int d^M \mathbf{q}_M \exp\{-\beta W_M(\mathbf{q}_M)\}. \quad (2)$$

\mathbf{q}_M is the complete set of microscopic coordinates necessary to describe a configuration of the system. $W_M(\mathbf{q}_M) = W(q_1, \dots, q_M)$ is the potential energy depending in general upon all coordinates. We imagine dividing configuration space into M -dimensional subvolumes ΔV_v (cells), which are so small that W_M can be assumed to have the constant value U_v within the cell ΔV_v . The probability P_v to find the system in a given small subvolume ΔV_v is

$$P_v = Z_M^{-1} \Delta V_v \exp\{-\beta U_v\}, \\ Z_M = \sum_v \Delta V_v \exp\{-\beta U_v\}. \quad (3)$$

The probability to find the system in an M -dimensional region A of configuration space is

$$P_A = Z_M^{-1} \left(\sum_v \right)_A \Delta V_v \exp\{-\beta U_v\}. \quad (4)$$

Here $(\sum_v)_A$ means that the sum is taken over all ΔV_v situated in the region A.

We now divide all possible values of W_M occurring in a given region into small intervals ΔE_i around energy values E_i . As such, we write Eq. (4) as

$$P_A = Z_M^{-1} \sum_i \exp\{-\beta E_i\} \left(\sum_v \right)_{A,i} \Delta V_v. \quad (5)$$

Here $(\sum_v)_{A,i} \Delta V_v$ means the sum over all those v for which the subvolume ΔV_v is situated in A and for which $E_i - \Delta E_i/2 < U_v \leq E_i + \Delta E_i/2$. The result of this sum is a volume that we call $f_A(E_i)$. Hence,

$$f_A(E_i) = \left(\sum_v \right)_{A,i} \Delta V_v. \quad (6)$$

Then Eq. (5) can be written as

$$P_A = Z_M^{-1} \sum_i f_A(E_i) \exp\{-\beta E_i\}. \quad (7)$$

P_A and Z_M both change with temperature, but $f_A(E_i)$ does not change with temperature because it is only the sum of subvolumes that belong to a given potential energy shell. Clearly $f_A(E_i)$ is the microcanonical partition function associated with the region A. The quantities

$$P_{A,i}(T^*) = Z_M^{-1}(T^*) f_A(E_i) \exp\{-\beta^* E_i\},$$

$$P_A(T^*) = \sum_i P_{A,i}(T^*) \quad (8)$$

can be directly evaluated from MC or MD simulations at a high temperature T^* , obviating the need to divide phase space into small subvolumes ΔV_v . $P_{A,i}(T^*)$ is the conditional probability to find the potential energy in the i th interval $[E_i - \Delta E_i/2, E_i + \Delta E_i/2]$ when the system is in state A.

Let us now assume that there is a second region B where the same procedure can be applied. Depending upon the application of HTCE (for TST, or, calculation of free energy differences or local entropy distributions or others) each one of state A or B can be identified with the reactant space or product space or transition state or any other region of the configuration space of the system. Further, we assume that N configurations of the system are ergodically sampled during a simulation at temperature T^* . Let $N_{A,i}$ be the number of events for which the system is found within the region A and with a potential energy W_M in the i th energy interval; analogously let $N_{B,i}$ be the corresponding quantity for region B. Then the probabilities $P_{A,i}$ and $P_{B,i}$ can be evaluated as

$$P_{A,i}(T^*) = \frac{N_{A,i}}{N}, \quad P_{B,i}(T^*) = \frac{N_{B,i}}{N}. \quad (9)$$

Hence, from Eq. (8) it follows that

$$f_A(E_i) = Z_M(T^*) \frac{N_{A,i}}{N} \exp\{\beta^* E_i\},$$

$$f_B(E_i) = Z_M(T^*) \frac{N_{B,i}}{N} \exp\{\beta^* E_i\}. \quad (10)$$

Because we do not generally know $Z_M(T^*)$, the individual functions $f_A(E_i)$ and $f_B(E_i)$ are unknown. However, the ratio P_B/P_A at any temperature T can be obtained according to:

$$\frac{P_B(T)}{P_A(T)} = \frac{\sum_i f_B(E_i) \exp\{-\beta E_i\}}{\sum_{i'} f_A(E_{i'}) \exp\{-\beta E_{i'}\}} = \frac{\sum_i N_{B,i} \exp\{(\beta^* - \beta) E_i\}}{\sum_{i'} N_{A,i'} \exp\{(\beta^* - \beta) E_{i'}\}}. \quad (11)$$

The unknown factor $Z_M(T^*)$ which is common to all f_A and f_B has cancelled and plays no role. Eq. (11) explicitly shows how statistics gathered at high temperature ($N_{A,i}$ and $N_{B,i}$ at T^*) can be used to construct a ratio of probabilities at other temperatures (T). Identifying state A with the reactant state and B with the transition state hence, P_B with $P_{\#}(\delta)$, completes the theory behind using HTCE for TST calculations.

3. Model system and results

We have tested the HTCE method on a simple model system for which direct MD rate calculations are possible, for comparison purposes. Two extended regions of low potential energy are separated by a threshold like two interconnected pores in a solid, e.g. a zeolite that are separated by a narrow 'window' (a bottleneck) that forms an energetic or entropic barrier. This is the typical pattern always connected with zeolites in which guest diffusion is too slow to be observed by MD simulations. We consider a single particle in the potential

$$W(x, y, z) = A_x x^4 + A_y y^2 + A_z z^2 + x(C_y y + C_z z). \quad (12)$$

The thermostat of Kast et al. [18,19] is used in order to establish coupling of the particle to a heat bath, thus ensuring the statistics of the canonical ensemble. The potential parameters are taken to be: $A_x = 0.005$, $A_y = 20.0$, $A_z = 20.0$, $C_y = 3.35$, $C_z = 2.55$. Fig. 1 shows the potential landscape with all coordinates in Å and energies in kJ/mol and hence, the parameters A_x , A_y , A_z , C_y and, C_z in kJ/(mol Å²).

The lowest potential energy is -39.272 kJ/mol at the following two minima: $(-9.414, 3.154, 2.401)$ and $(9.414, -3.154, -2.401)$. The saddle point (transition state) is at $(0, 0, 0)$ with zero potential energy.

We have sampled this potential using unconstrained, thermostatted molecular dynamics. In the real system the thermal motion of the atoms of the solid would provide the thermostat.

The mass of the lone particle is taken to be 16 g/mol representing, e.g., the mass of methane. The equations of motion are integrated with the velocity Verlet algorithm modified by momentum changes corresponding to collisions with imaginary bath particles (with masses of 0.03 g/mol) according to the algorithm of Kast et al. [18,19].

The brute-force MD simulations that determine the 'exact' rate constants were run at 700–1000 K for 2×10^9 – 5×10^9 time steps (total times of 2–5 μs), depending on the temperature. The HTCE sampling simulations were run at four temperatures roughly in the range 650–4000 K, to determine how hot the system needs to be to provide sufficient statistics. The HTCE sampling at 3986 K was run for various total times in the range 1–100 ns, to determine the total times needed for statistical convergence (see below).

The transition state and the two minima are connected in three-dimensions by a straight line because of the bilin-

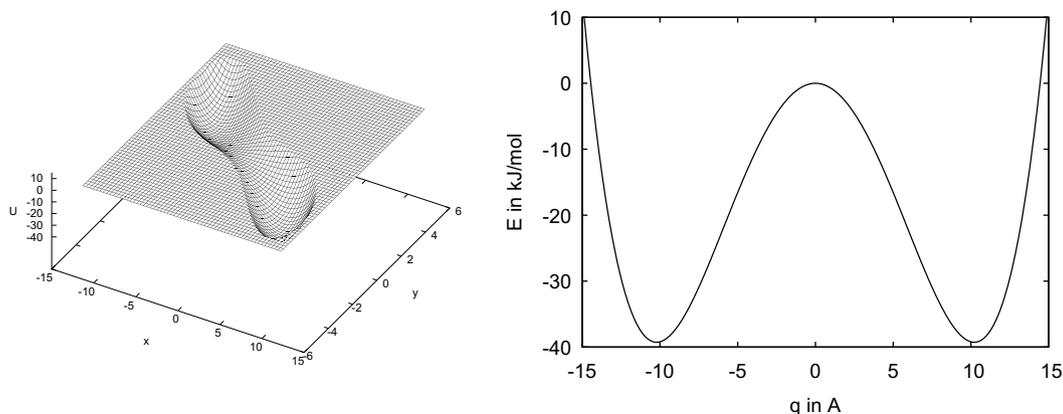


Fig. 1. Potential landscape for the simple model (energy in kJ/mol; distances in Å). Left: potential energy along reaction coordinate. Right: Potential energy in the xy -plane up to 10 kJ/mol.

ear coupling in Eq. (12). We take this line to be the reaction coordinate q measured in Å. We thus define the reactant region A by $q < -\delta/2$, and the transition state region # by $q \in [-\delta/2, \delta/2]$. Because the model potential changes slowly in the transition region, we found good agreement with brute-force MD simulations of the rate for $\delta = 1$ Å. For histogramming the potential energy, we have used a constant $\Delta E = 1$ kJ/mol.

Fig. 2 shows normalized probability densities $P_A(E)$ and $P_{\#}(E)$ to find the system with a given potential energy E in reactant state A and at the transition state #. The normalization has been chosen according to Eq. (8) in such a way that integrating $P_A(E)$ over all energies E yields P_A . The functions $P_{\#}(E)$ are multiplied by the same factor. Note, however, that the normalization cancels out in all ratios; this normalization has been applied only for demonstration purposes in Fig. 2.

For our symmetric model system $P_A = 0.5$; this is fulfilled with good accuracy for temperatures higher than about 700 K. However, for 647 K, even with a run duration of 2 μ s, ergodicity was not yet achieved. The system spent 10.8 times longer in the product space than in the reactant space, leading to $P_A(E)$ values that are much too

small. The low MD rate coefficient from this run is another hint that MD at this temperature (and at lower ones) is not suited to calculate the $P_A(E)$ and $P_{\#}(E)$ distributions, nor to evaluate the rate from brute-force MD.

Distributions are shown from HTCE sampling at 647, 996, 1997 and 3986 K. These distributions naturally broaden to higher energies with increasing temperature.

Fig. 3 shows rate coefficients computed using brute-force MD for temperatures in the range 700–1000 K. At lower temperatures these brute-force MD rate coefficients are too inaccurate to provide meaningful comparisons, even for MD runs as long as 5 μ s (5×10^9 time steps). For example, at 648 K, forward and backward rate constants differed by more than an order of magnitude despite the symmetry of the Hamiltonian. Also shown for comparison in Fig. 3 are TST rate constants computed using Voter's displacement vector Monte Carlo method [12]. Of course, the dynamically exact brute-force MD rate constants and TST rates should not agree perfectly. However, the classical transmission coefficient (dynamical correction factor) is but a small effect in Fig. 3, which covers seven orders of magnitude in time scales. The agreement between these two sets of reference rate coefficients is excellent.

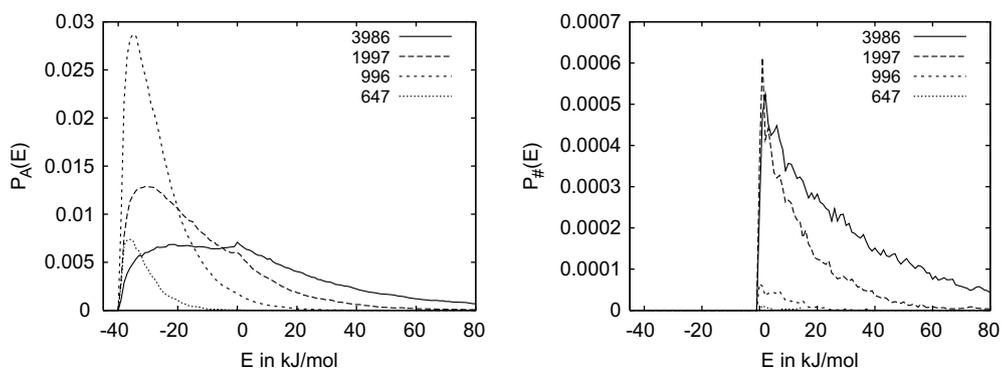


Fig. 2. Normalized (see text) probability density $P(E)$ to find a given potential energy E , as a function of temperature (which is given in K). Left: reactant state A; right: transition state #.

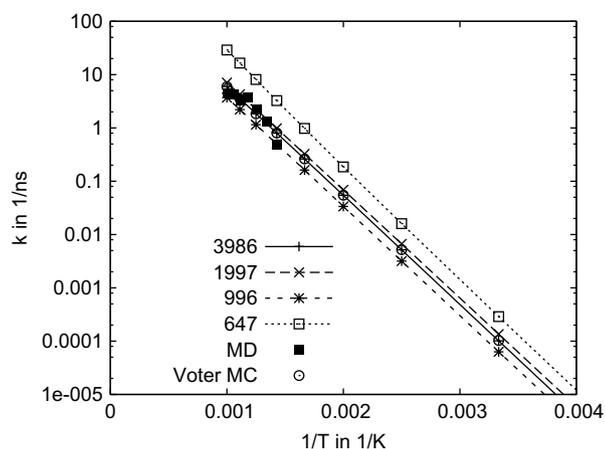


Fig. 3. Arrhenius plot of HTCE/TST rate constants sampled at four temperatures (in K), compared with brute-force MD results, and with TST rate constants from Voter's displacement vector method.

Fig. 3 also shows rate coefficients obtained from HTCE sampling according to Eqs. (1) and (11). The comparison shows that the HTCE/TST rates computed from statistics obtained at 3986 K give quantitative agreement with both brute-force MD and Voter's Monte Carlo TST rates, all the way down to 300 K. HTCE rates computed from sampling at 1997 K give very good but not excellent agreement with the 'exact' rates. HTCE rate constants obtained from sampling at even lower temperatures give progressively worse agreement. At lower temperatures the high energy regions, particularly the transition region, are not frequently visited and the statistics in exploring the configuration space is too poor for these regions. This means no ergodicity (as mentioned above).

The slope extracted from HTCE rates (3986 K) in Fig. 3 gives an apparent activation energy of 39.20 kJ/mol, in excellent agreement with the classical barrier height of 39.27 kJ/mol. This confirms the effectiveness of the HTCE approach for gathering statistics at one high temperature for use at many lower temperatures.

4. Convergence of the HTCE method

Here we explore the convergence of the HTCE method with respect to simulation time for high-temperature sampling. As a reference, we focus on the TST partition function ratio at 700 K. In Fig. 4, we show the convergence of both Voter's Monte Carlo and the new HTCE method. All ratios in Fig. 4 are divided by the Voter result from the longest MC run (10^8 MC steps); as such, all results converge to unity. The efficiency of Voter's method was enhanced by using a boost potential of $V = 39$ kJ/mol as described in Ref. [12]. This boost essentially eliminates the large potential energy difference between reactant state A and the transition state #. The HTCE rates were computed using statistics gathered from MD at 3986 K. The results in Fig. 4 show that 2% accuracy in the partition function ratio

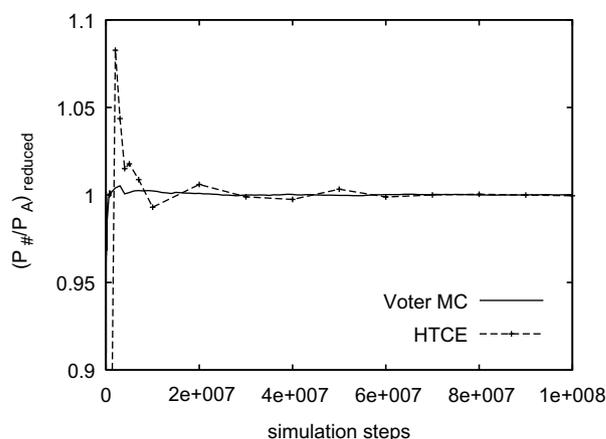


Fig. 4. Convergence of the partition function ratio at 700 K, comparing HTCE and Voter's Monte Carlo method [12], as a function of simulation length. All values have been divided by the final value of Voter's method.

is obtained using Voter's method after 10^5 MC steps, while the same accuracy is obtained from HTCE after 4×10^6 MD steps. Although HTCE requires a longer run, HTCE provides information that allows TST rate calculations over a wide range of temperatures, whereas the Voter method requires two separate MC simulations for each temperature. As such, the HTCE method becomes more efficient in this case if TST rates are required for 40 or more temperatures, e.g. to examine if an Arrhenius law is valid.

5. Concluding remarks

We have reported a new approach for computing partition function ratios, with application to TST rate coefficients. The method samples configuration space with one high-temperature Monte Carlo (MC) or molecular dynamics (MD) run. Potential energy histograms are accumulated during the high-temperature run, from which microcanonical factors can be extracted as functions of potential energy. Canonical partition functions can then be constructed at much lower temperatures. The method was demonstrated on a model system for which brute-force exact calculations are possible. Excellent agreement was obtained for activated crossing rates of a 40 kJ/mol barrier, for temperatures as low as 300 K. This new approach, which we call high-temperature configuration-space exploration (HTCE), becomes the method of choice when TST rates at many temperatures are required. At present we got promising results from an investigation of pentane in the zeolite ZK5. Because of the complicated nature of this system in which also changes of the conformation of pentane appear we have compared the HTCE results with results from transition path sampling. These investigations will be presented in a forthcoming full paper [21].

One drawback of HTCE is that, to gather sufficient statistics, the dynamics are run at temperatures that may be outside the physical regime of the material under study,

e.g., above the melting temperature of a zeolite. We imagine several solutions to this problem. In the first instance, one may consider fixing some coordinates, thermostating the remaining (interesting) degrees of freedom, and compiling HTCE statistics as shown above.

The problem with this approach is that solid vibrations that may impact barrier crossing are ignored. To take such vibrations into account, one may consider replacing anharmonic solid–solid potentials with harmonic ones, allowing solid vibrations to be sampled at arbitrarily high temperatures. The disintegration of the real system that would appear at high temperatures is not the subject of this investigation. Therefore, harmonic intramolecular potentials for the solid and the guest molecule as well that yield the vibration frequencies at moderate temperatures satisfactorily are sufficient. Doing this in the above mentioned system of pentane in ZK5 we found agreement with experimental values for the self-diffusion coefficient at room temperature [21].

The main virtue of the HTCE method is the simplicity of its implementation: partition function ratios can be computed without special importance sampling. In addition, any quantity that can be expressed as a partition function ratio, e.g., free energy and entropy differences [20], can be computed using HTCE. This method is generally applicable even for systems that exhibit non-Arrhenius temperature dependencies. We are applying this to more realistic diffusion systems [21].

Acknowledgements

SF and AS thank the DFG for financial support within the framework of the SPP1155 Project FR1486/2-2. SF and AS gratefully acknowledge the Computer Center of the University of Leipzig for access to facilities and for computer time. We thank Horst Vörtler (Leipzig) for fruitful discussions.

References

- [1] D.G. Truhlar, B.C. Garrett, S.J. Klippenstein, *J. Phys. Chem.* 100 (1996) 12771.
- [2] F. Keil, R. Krishna, M.-O. Coppens, *Rev. Chem. Eng.* 16 (2000) 71.
- [3] R. Haberlandt, S. Fritzsche, H.L. Vörtler, Simulation of microporous systems: confined fluids in equilibrium and diffusion in zeolites, in: H.S. Nalwa (Ed.), *Handbook of Surfaces and Interfaces of Materials*, vol. 5, Academic Press, San Diego, London, Boston, New York, Sidney, Tokyo, Toronto, 2001, pp. 358–444.
- [4] J. Kärger, S. Vasenkov, S.M. Auerbach, Diffusion in zeolites, in: S.M. Auerbach, K.A. Carrado, P.K. Dutta (Eds.), *Handbook of Zeolite Science and Technology*, Marcel Dekker Inc., New York, 2003, pp. 341–422.
- [5] R.Q. Snurr, A.T. Bell, D.N. Theodorou, *J. Phys. Chem.* 98 (1994) 5111.
- [6] H. Ramanan, S.M. Auerbach, Modeling jump diffusion in zeolites: I. Principles and methods, in: J. Fraissard, W.C. Conner (Eds.), *NATO-ASI Series C: Fluid Transport in Nanopores*, Kluwer Academic Publishers, Dordrecht, 2004.
- [7] M. Mella, *J. Chem. Phys.* 122 (2005) 204106.
- [8] A.F. Voter, *Phys. Rev. Lett.* 78 (1997) 3908.
- [9] C. Dellago, P.G. Bolhuis, P.L. Geissler, *Adv. Chem. Phys.* 123 (2002) 1.
- [10] M.R. Sorensen, A.F. Voter, *J. Chem. Phys.* 112 (2000) 9599.
- [11] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, Academic Press, San Diego, London, Boston, New York, Sidney, Tokyo, Toronto, 2002.
- [12] A.F. Voter, *J. Chem. Phys.* 82 (1985) 1890.
- [13] F. Jousse, S.M. Auerbach, *J. Chem. Phys.* 107 (1997) 9629.
- [14] See, e.g., Keith Refson, <<http://www.earth.ox.ac.uk/~keithr/moldy.html>>.
- [15] T. Mosell, G. Schrimpf, J. Brickmann, *J. Phys. Chem.* 101 (1997) 9476.
- [16] T.R. Forester, W. Smith, *Faraday Trans.* 93 (1997) 3249.
- [17] A.M. Ferrenberg, R.H. Swendsen, *Phys. Rev. Lett.* 63 (1989) 1195.
- [18] S.M. Kast, K. Nicklas, H.-J. Bär, J. Brickmann, *J. Chem. Phys.* 100 (1994) 566.
- [19] S.M. Kast, J. Brickmann, *J. Chem. Phys.* 104 (1996) 3732.
- [20] A. Schüring, S.M. Auerbach, S. Fritzsche, R. Haberlandt, *J. Chem. Phys.* 116 (2002) 10890.
- [21] O. Saengsawang et al. Self-diffusion of pentane in the zeolite ZK5 investigated by transition state theory, in preparation.