New method to analyze simulations of activated processes

Jan Wedekind[a] and Reinhard Strey
Institut für Physikalische Chemie, Universität zu Köln, D-50939 Köln, Germany

David Reguera
Departament de Física Fonamental, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain

(Received 27 November 2006; accepted 9 February 2007; published online 2 April 2007)

We present a new method to analyze molecular and Brownian dynamics simulations of activated processes based on the concept of mean first-passage times. The new method provides a simple and efficient strategy to evaluate reaction rates and it facilitates the localization of the transition state directly from the kinetics of the system without the need of thermodynamical considerations. It also provides a more rigorous value of the steady-state transition rate and gives valuable information about many important characteristics of the process. We illustrate the power of this new technique by its application to the study of nucleation in rare gases. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2713401]

I. INTRODUCTION

Many important processes in nature are activated. Examples embrace chemical reactions, polymer and protein conformational changes, nucleation, earthquake formation, crystallization, cavitation, or the formation of cracks or bubbles, thus spanning a wide variety of different phenomena and fields. These kinds of processes are characterized by the presence of a free energy barrier that has to be overcome by the occurrence of a rare fluctuation in order to change the system from one state to another. A proper understanding of the kinetics of these processes is thus the cornerstone of many disciplines and a very challenging task.

A promising route of investigating activated processes is via computer simulations. Molecular dynamics (MD) simulations are particularly suited for this task since they allow us to track the real dynamics of the process at the molecular scale. Brownian dynamics simulations also provide a useful technique to investigate the dynamics of these systems at larger time scales. However, the difficulty often lies in the analysis of the simulation results. The crossing of a free energy barrier that is characteristic of an activated process is a rare event and consequently difficult to observe and quantify in a simulation. In addition, it is a stochastic process: it may occur after a short or a long time, or not occur at all in the limited time of a simulation. All this is further complicated by the fact that in most cases the location of the transition state, i.e., the maximum of the barrier, is not known *a priori*.

Here, we propose a new method to analyze the kinetics of activated processes in simulations based on the concept of mean first-passage times (MFPT). This new method has many advantages: it is very easy and straightforward to implement in a simulation and not only allows a precise determination of the rate but in addition facilitates the location of the transition state purely from the kinetics. Moreover, it allows us to differentiate the activation from the subsequent evolution or growth of the system and the possibility to infer more valuable information such as the activation barrier, growth curves, and lag times. Since these important quantities are evaluated from the kinetics alone, these kinetic estimates can then be compared with thermodynamical (or equilibrium) predictions.

The article is set up as follows. In Sec. II, we introduce the concept of mean first-passage time and describe the theoretical basis of the new method. In Sec. III, we illustrate the usefulness of the method with one of the prototypical examples of an activated process: vapor-liquid nucleation. We first verify the accuracy of the new method against a simple analytical model, the classical nucleation theory, and then, we describe the practical implementation in a MD simulation using the condensation of Lennard-Jones argon as an example. The main conclusions are finally summarized in Sec. IV.

II. THEORETICAL DERIVATION

The dynamics of most activated processes can be recast into a Kramers-type description in terms of a reaction coordinate x and described by a Fokker-Planck equation[1]

$$\frac{\partial P(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(D_0 e^{-\beta U(x)} \frac{\partial}{\partial x}(P(x,t)e^{\beta U(x)}) \right), \tag{1}$$

where $P(x,t)$ is the probability distribution function, D_0 the generalized diffusion coefficient, $U(x)$ the free energy barrier that has to be overcome, and $\beta = 1/kT$, with k the Boltzmann constant and T the temperature. The reaction coordinate x can be the size of a cluster or an aggregate, the angle of a bond, or in general any variable that properly describes the state of the system.

In an activated process, the quantity of major interest is the rate at which the barrier is crossed. For this type of dynamics, the steady-state rate of barrier crossing can be related to the MFPT.[2] The MFPT $\tau(x_0;a,b)$ is defined in the one dimensional case as the average elapsed time until the system starting out at point x_0 leaves a prescribed domain $[a,b]$ for the first time [see illustration in Fig. 1(a)]. In gen-

[a]Electronic mail: jan.wedekind@uni-koeln.de

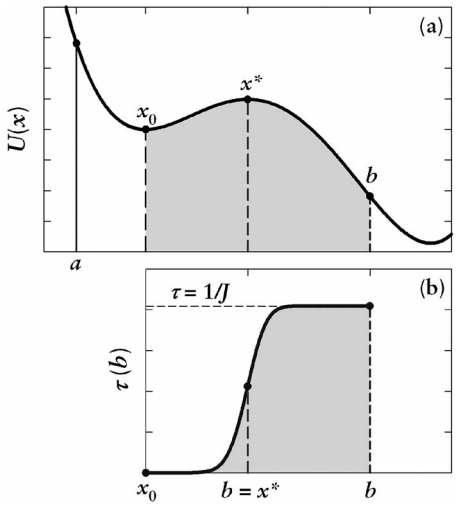

FIG. 1. Schematic representation of an activated process and its corresponding MFPT (for details, see the text).

eral, the MFPT depends on the nature of the boundary conditions and on the initial position x_0. For most cases of interest, the boundary conditions are typically reflecting at point a and absorbing at point b. Under these circumstances, the MFPT for a process described by Eq. (1) is given by the expression[2,3]

$$\tau(x_0;a,b) = \int_{x_0}^{b} \frac{1}{D_0} dy \exp[\beta U(y)] \int_{a}^{y} dz \exp[-\beta U(z)]. \tag{2}$$

Typically, this MFPT is analyzed in terms of the starting position x_0. Let us now assume that we fix a given starting point and look at the behavior of the MFPT in terms of b, from now on denoted by $\tau(b)$ [see Fig. 1(b)]. That will give us information about the average time it takes the system to reach b for the first time. For instance, $\tau(b=x^*)$ is the average time needed to reach the transition state x^*, i.e., the top of the barrier, for the first time. It is well known that the barrier-crossing rate is related to that time by[2,4]

$$J = \frac{1}{2\tau(x^*)}, \tag{3}$$

where the factor $1/2$ comes from the fact that, at the top of the barrier, the system has a 50% chance of falling to either side.

Besides the activation rate, we can also get a precise estimation of the location of the transition state x^* from the behavior of the MFPT. For the case of a constant diffusion coefficient D_0, the first derivate of the MFPT [Eq. (2)] with respect to b yields

$$\frac{\partial \tau(b)}{\partial b} = \frac{1}{D_0} \exp[\beta U(b)] \int_{a}^{b} dz \exp[-\beta U(z)]. \tag{4}$$

Taking the second derivate we get

$$\frac{\partial^2 \tau(b)}{\partial b^2} = \frac{1}{D_0} + \beta U'(b) \frac{\partial \tau(b)}{\partial b}, \tag{5}$$

which—evaluated at the transition state x^*, i.e., at the top of the barrier where $\beta U'(x^*)=0$—finally yields

$$\left.\frac{\partial^2 \tau(b)}{\partial b^2}\right|_{b=x^*} = \frac{1}{D_0}. \tag{6}$$

This equation provides a simple and unique criterion to determine the location of the transition state x^* merely from the kinetics. Reversely, in situations where the location of the transition state is already known or can be obtained from other methods, Eq. (6) will serve to determine the kinetic prefactor D_0 in a very simple way.

It is worth stressing that activation rates are only properly defined when the barrier is relatively high, i.e., $\beta U(x^*) \gg 1$. This assures a proper separation between the time scale associated with the barrier-crossing event and that of the local diffusion or relaxation. Under those circumstances, the MFPT shows a characteristic sigmoidal shape [see Fig. 1(b)], and the rate is simply given by the inverse of the value of the MFPT at the plateau, and relatively insensitive to the initial condition x_0 (for $x_0 < x^*$) and to the location of b (for $b > x^*$). Moreover, the MFPTs are then significantly bigger than the characteristic diffusion time $1/D_0$. Therefore, under these circumstances, Eq. (6) simplifies to

$$\left.\frac{\partial^2 \tau(b)}{\partial b^2}\right|_{b=x^*} \cong 0. \tag{7}$$

Thus, the inflection point of the MFPT curve indicates x^* (the top of the barrier).

Yet, a more powerful result can be derived, which is the basis of our new method. Under reasonably high barriers, the behavior of the MFPT, Eq. (2), in the vicinity of the critical size x^* can be evaluated by the method of steepest descent (see the Appendix) and it is given by the following expression:

$$\tau(b) = \frac{\tau_J}{2}(1 + \mathrm{erf}((b - x^*)c)), \tag{8}$$

where x^* is again the location of the transition state, $\mathrm{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{-x^2} dx$ is the error function,

$$c = \sqrt{\frac{|U''(x^*)|}{2kT}} \tag{9}$$

is the local curvature around the top of the barrier, and

$$\tau_J = \frac{1}{J} \tag{10}$$

is the inverse of the (steady-state) transition rate. Equation (8) offers a very powerful method to analyze the results of simulations of activated processes. By evaluating the MFPT in the simulation and by merely fitting the results to this simple expression, we can immediately get all the important parameters of the process. Namely, we obtain the barrier-crossing rate J, the location of the transition state x^*, and the curvature at the top of the barrier $c = \sqrt{|U''(x^*)|/2kT}$. This

parameter c is also related to the time lag to reach steady-state conditions for the rate, which is (except for a numerical factor) given by $\tau_{\mathrm{lag}} \sim 1/(D_0 c^2)$.[5]

There is an intimate relation of Eq. (8) with a similar expression obtained in the context of nucleation from the growth probability of clusters.[6,7] In a pioneering work, ter Horst and Kashchiev[6] derived an expression to determine the critical cluster size n^* and the so-called Zeldovich factor[4] $Z=c/\sqrt{\pi}$ from the growth probability of clusters in one-component nucleation. Later, they also showed how the nucleation rate J can be obtained from the growth probabilities of dimers.[8] Their method requires the measurement of the growth probability $P(n)$, defined as the probability that a cluster of size n grows up to a sufficiently large supercritical size $M \gg n^*$. In practice, this probability can be evaluated as $P(n)=N/N_0$, where N is the number of experiments in which the cluster has reached this size, and N_0 is the total number of experiments. Moreover, this growth probability can be approximated by[6]

$$P(n) = \tfrac{1}{2}(1 + \mathrm{erf}((n - n^*)c)). \tag{11}$$

When the barrier is high enough, $P(n)$ can also be approximated by the ratio of the MFPT $\tau(n)$ of an n-sized cluster and the MFPT of a sufficiently large supercritical cluster τ_J,

$$P(n) = \frac{\tau(n)}{\tau_J}. \tag{12}$$

We can immediately see that a combination of Eqs. (11) and (12) yields the same result as Eq. (8). Therefore, in the high barrier case Eqs. (11) and (8) provide two alternative but equivalent ways to analyze the cluster growth history. The main difference between both approaches lies on the practical implementation in different simulation techniques. It is simpler to obtain first-passage *times* from a molecular or Brownian dynamics simulation than growth probabilities of clusters. In addition, and contrary to the growth probability, the MFPT method provides full information about the dynamics of the process and, as we will see, about the possible coupling between nucleation and growth. Growth probabilities, on the other hand, can be relatively easily calculated from stochastic simulations (provided the growth and decay rates are known) or from a Monte Carlo simulation, where different n clusters can be prepared independently and their evolution monitored. Therefore, each method is particularly suited to different simulation techniques and, in conjunction, both methods provide invaluably powerful tools to analyze activated processes by practically any available simulation technique.

Finally, Bartell and Wu recently presented a similar result as Eq. (8) in the context of freezing.[9] We would like to clarify that we have already presented the MFPT method to the community in several conference talks prior to the submission of Ref. 9.[10,11] In addition, it was also derived and extensively used in a Ph.D. thesis,[12] part of which has already been published,[13] prior to the publication of Ref. 9. We find that neither a rigorous derivation nor a sufficient justification is given in Ref. 9. Moreover, the applicability of the result of Ref. 9 is not demonstrated analytically or by using a suitable set of simulation results. Instead, only four data points from old simulation results are used, which in turn had to be obtained from an interpolation in order to gain at least some kind of estimate of the first appearance times (unfortunately, no error bars are presented in the respective figure). Here, we present a full derivation and justification of the MFPT method and in the following we will showcase the power and applicability of the MFPT both analytically and using MD simulations for a prototypical case of an activated process.

III. A CASE STUDY ON VAPOR-LIQUID NUCLEATION

For the sake of concreteness, let us now focus on one particular example of an activated process, namely, vapor-liquid nucleation. We will first validate the method by using an analytical model, the classical nucleation theory (CNT), and then illustrate its application to simulation results.

A. Analytical comparison with CNT

Nucleation is the mechanism that initiates many phase transformations. For instance, the condensation of a vapor starts through the formation of small liquid droplets that have to reach a certain critical size to be able to grow and complete the phase transformation. The activated nature of this process arises from the energetic cost, related to the surface tension, that has to be payed to build these small clusters in the supersaturated vapor.

To describe vapor-liquid nucleation, a proper reaction coordinate is the size n, or the number of molecules in the liquid cluster, and $U(x)$ would represent the free energy of formation of a cluster of n molecules. In CNT, this free energy is given by

$$U(x) \equiv \Delta G(n) = -n\Delta\mu + \sigma A(n), \tag{13}$$

where $\Delta\mu$ is the difference in chemical potentials of the liquid and the vapor phase, σ is the surface tension, and $A(n)$ is the area of a spherical cluster of n molecules. One can show that this free energy can be rescaled to[14]

$$\Delta G(x) = 2\Delta G^* \left(-x + \tfrac{3}{2}x^{2/3}\right), \tag{14}$$

where $x=n/n^*, n^*$ is the size of the critical cluster, and ΔG^* is the height of the nucleation barrier. This representation has the advantage that the location of the barrier is easily identified at $x=1$ independent of the barrier height. Figure 2(a) shows two examples of the rescaled free energy, Eq. (14), for $\Delta G^*=5kT$ and $20kT$. It is well known that the nucleation kinetics can be described by the Fokker-Planck Eq. (1).[15] In this case, the effective diffusion D_0 is the rate of attachment of molecules to a cluster, which is reasonably accurately given by kinetic theory as $D_0(x)=\beta_0 x^{2/3}$, with $\beta_0=A(n^*)p/\sqrt{2\pi mkT}$, and p the vapor pressure.

From Eqs. (2) and (14) we can now calculate the corresponding MFPTs analytically (using a reflecting boundary condition at $x=0$ and $x_0=0$). Figure 2(b) illustrates the resulting MFPT, in units of n^{*2}/D_0, as a function of the cluster size for the two cases in Fig. 2(a), corresponding to a high and a low barrier, respectively. For the high barrier, the

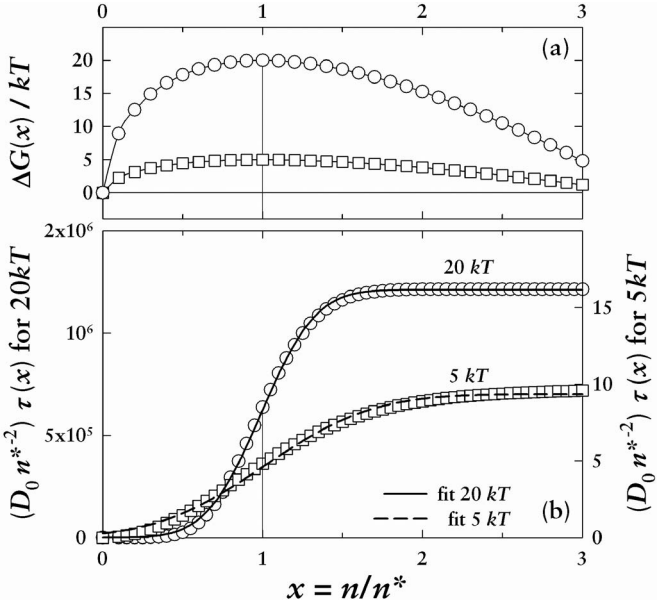

FIG. 2. (a) Two free energy barriers of 5 and $20kT$ [Eq. (14)]. (b) Corresponding MFPTs (symbols) calculated from Eq. (2) and their fit to Eq. (8).

MFPT curve has a sigmoidal shape that reaches a well-defined plateau at larger values of x. The presence of this plateau indicates that the rate limiting step in the formation of a large cluster is the activation time to overcome the critical size and that the time required for the subsequent growth of the cluster is negligible compared to that. For the low barrier case, the MFPT still increases noticeably for higher n without reaching a clear plateau, indicating that the growth of the cluster occurs at similar time scale as its nucleation. It is also very important to notice the huge difference of five orders of magnitude in time scales in Fig. 2(b), which reflects the extreme sensitivity of the passage times and the rates to the height of the nucleation barrier.

Let us now see how accurately the fitting function Eq. (8) can represent the results of the analytically calculated MFPT. Figure 2(b) shows the results of a simple least squares fitting to Eq. (8).[17] The parameters of the fit, listed in Table I, reproduce the exact values of the nucleation rate $J=1/\tau_J$, the critical size x^*, and the so-called Zeldovich factor,[4] defined as $Z\equiv\sqrt{|U''(x^*)|/2\pi kT}=c/\sqrt{\pi}$, with excellent accuracy in both cases. This demonstrates that the fitting procedure to Eq. (8) works remarkably well even for the case of a very low barrier, where the steepest descent approximation is expected to fail and the rates are not well defined.

Once having verified the accuracy and validity of the method, we will now discuss its application to simulation in the next subsection.

B. Application in MD simulations

The importance and high appeal of the new method lie in its practical implementation in a simulation. The evaluation of the nucleation rate directly from MD simulations is a difficult task. Direct or "brute-force" MD simulations[16]—where one simply starts from a supersaturated vapor and just waits for the appearance of critically sized clusters—are limited to situations in which the activation barrier is not very high. If the activation barrier is too high, the spontaneous crossing of the barrier becomes very unlikely and cannot be observed in the limited time accessible in simulation.[17] In such cases, one has to resort to indirect and more complicated techniques such as umbrella sampling[18] or transition path sampling.[19] Still, brute-force MD simulations offer a direct, unbiased, and detailed description of the kinetics of activated systems for moderate activation barriers and they are extensively used to investigate nucleation, crystallization, and many other activated processes. Typically, in brute-force simulations, two different methods to estimate the rate have been used to date. Since the size of the critical cluster is not known *a priori*, the most common method involves choosing a big enough threshold size and counting the time required to form a cluster of that size. However, the choice of this threshold is necessarily arbitrary, which leads to an inaccurate estimate of the nucleation rate. Another method is to count how many large clusters are present in the system as a function of time.[20] However, it requires a very large system and is affected by significant depletion effects, which complicate the interpretation of the results. More elaborated ideas based on the analysis of the transient nucleation or the moments of the rate have been proposed recently.[21] Our new method offers significant advantages and a very simple and efficient way of extracting all relevant information through a simple fit to the MFPT.

To illustrate the power of the method, we have applied it in a real simulation of condensation of Lennard-Jones (LJ) argon vapor. We have performed MD simulations of $N=343$ Lennard-Jones argon atoms at $T=50$ K for two box sizes of $V=(16\text{ nm})^3$ and $(18\text{ nm})^3$, corresponding to a low and a high activation barrier, respectively. Details of the simulation will be provided elsewhere.[22] It is worth mentioning, however, that the small system size was determined to be sufficiently large without encountering significant finite-size effects in the simulations.[13] The use of a small system

TABLE I. Comparison of the exact values of the nucleation rate $J=1/\tau_J$, the critical size $x^*=n/n^*$, and the Zeldovich factor $Z=c/\sqrt{\pi}$, with the parameters obtained from the fit to Eq. (8).

	$\Delta G^*=20kT$			$\Delta G^*=5kT$		
	Exact	Fit	Err. (%)	Exact	Fit	Err. (%)
J	8.236×10^{-7}	8.299×10^{-7}	0.8	0.1036	0.1044	1.3
x^*	1	0.99	1.0	1	1.03	3.0
Z	1.46	1.47	1.2	0.728	0.715	1.6

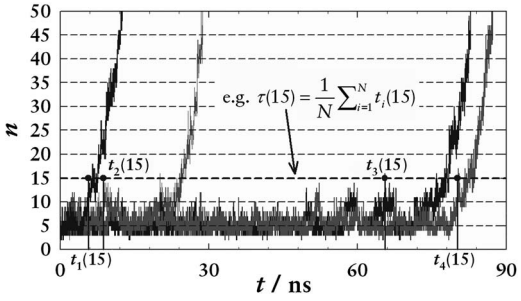

FIG. 3. Four example growth curves of the largest cluster from the MD simulations. For each run, we note the time it takes to reach a given size n for the first time. Averaging over all runs yields the MFPT $\tau(n)$ for each size n.

facilitates very good statistics by making it possible to repeat the simulations at the same conditions but with different starting configurations of the vapor. The same number of nucleation events can be observed much faster than in one large scale simulation because the computational effort strongly depends on the total number of molecules in the simulation.

The procedure to obtain the MFPT directly from the MD simulation is, in fact, very simple and illustrated in Fig. 3. For each simulation, the size of the largest cluster in the system is noted at regular intervals (1000 time steps in our case) and the time at which each size n appears for the first time, $t_i(n)$, is stored. The *mean* first-passage time $\tau(n)$ for each size n is simply obtained by averaging $t_i(n)$ over several repetitions of the simulations with different initial configurations. It should be noted that this method *does not* depend on whether we have just one or many clusters in the system or whether consecutive sizes are first reached by the same cluster or not. In our case, we performed 200 simulations at $V=(18\ \mathrm{nm})^3$ and 1000 simulations at $V=(16\ \mathrm{nm})^3$.

Figure 4 plots the resulting MFPTs obtained from the simulation and the corresponding fits to Eq. (8). The fit works very well for the case of $V=(18\ \mathrm{nm})^3$ and immediately provides all the relevant information from the simulation in a simple and accurate way. For the case of $V=(16\ \mathrm{nm})^3$ the MFPT does not reach a plateau but the fit is nevertheless reasonably good around the inflection point of the curve and provides a very good estimate of the critical size, which is well defined by Eq. (7). Again, it is an important advantage of the MFPT method is that we can clearly observe whether nucleation and growth are coupled, i.e., whether or not the barrier is significantly higher than kT. This condition obviously holds for the case of $V=(18\ \mathrm{nm})^3$, where the MFPT reaches a clear plateau at τ_J, and the nucleation rate simply is $J=1/(V\tau_J)$. However, this regime is typically not easily accessible in a brute-force MD simulation. For instance, in the $V=(16\ \mathrm{nm})^3$ case no plateau is reached, indicating that the barrier is low and nucleation and growth occur at the same time scale. Here, the standard threshold method to evaluate the rate would be inaccurate. It is important to realize that in such a case $J=1/(V\tau_J)$ only represents the rate of formation of a critical cluster. For low nucleation barriers, however, the overcoming of the barrier is no longer the rate limiting step and the rate of formation of an n-sized postcritical cluster depends both on J and on the growth time from size n^* to size n.

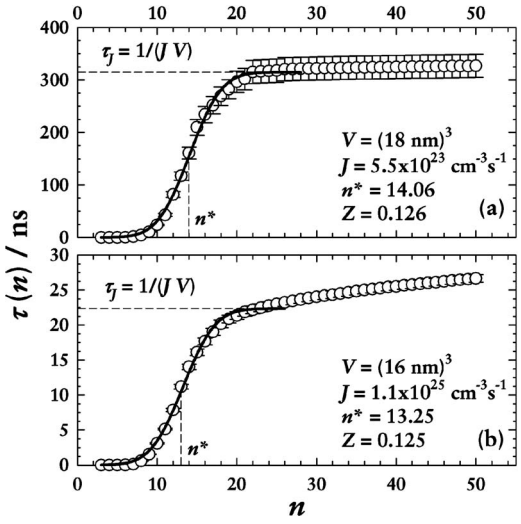

FIG. 4. Simulation results (circles) of LJ argon at 50 K and two different volumes corresponding to a low (a) and a high (b) barrier, respectively. 1000 simulations were performed for (b) and 200 for (a). Dashed lines are fits to Eq. (8).

On the other hand, the location of the barrier n^* is still accurately defined by Eq. (8), i.e., the turning point of the MFPT curve. Furthermore, despite the absence of a plateau, the MFPTs in the vicinities of the critical size are reasonably well described by the fit and, as verified in Sec. III A, the rate, critical size, and Zeldovich factor may still be accurately determined from Eq. (8). Another possible approach to determine n^* would be to sample the growth probability $P(n)$ and accurately fit them to Eq. (11), since $P(n)$ will always reach a plateau.[6] Yet, it is straightforward to show that such a procedure is fully equivalent to using Eq. (7), i.e., finding the turning point of the MFPT. However, the growth probability $P(n)$ method will not provide any accurate estimate of rates nor a simple way to distinguish between steady-state nucleation from nucleation/growth coupling, which can be easily identified in Fig. 4.

In addition, the MFPT method provides a nice and clear way to contrast thermodynamic and kinetically determined parameters. For instance, we can test the accuracy of the different cluster criteria (or reaction coordinates in general) by comparing the size of the critical cluster obtained kinetically with independent thermodynamic estimates. In our case, a simple Stillinger criterion was used to characterize the number of molecules of the clusters in the simulation. In this criterion, two atoms that are closer than a certain threshold distance r_S, the Stillinger radius, are considered to form part of a cluster. Obviously, the cluster size determined in the simulation, and, in particular, the critical cluster size, will depend on this definition, and ideally, the optimal cluster definition should be the one providing a good measure of the actual size of the cluster.

In order to investigate how the cluster definition (or in general, the choice of the reaction coordinate) influences the results obtained from the MFPT, the same systems of Fig. 4 were simulated using two different Stillinger radii, one at $r_S=1.8\sigma_{\mathrm{LJ}}$ and one at $r_S=1.2\sigma_{\mathrm{LJ}}$. The results of the simula-

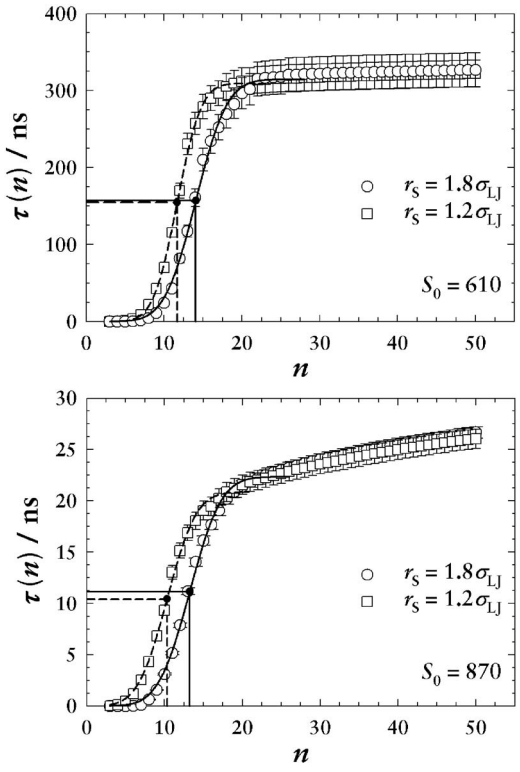

FIG. 5. Comparison of simulation results for different Stillinger radii r_S for the same systems as in Fig. 4, i.e., (a) $V=(18\ \mathrm{nm})^3$ and (b) $V=(16\ \mathrm{nm})^3$. The different cluster definition changes the obtained value for n^* but has only a minimal influence on the determined nucleation rate (8).

tions are shown in Fig. 5. By fitting the MFPT to Eq. (8) we can accurately determine the critical cluster size n^* (and also the nucleation rate) as measured by the two different cluster definitions. For the smaller Stillinger radius, the measured critical cluster size is slightly smaller than for the larger radius, while the nucleation time is practically unaffected. Thus, the nucleation rate determined by the MFPT method is independent of the particular cluster definition used in the simulation, confirming a similar conclusion of ter Horst and Kashchiev.[6] Table II shows the exact results obtained from the different fits in Fig. 5.

We can now use an independent thermodynamic estimate of the critical cluster size, for instance, the nucleation theorem, to decide which cluster criteria provide the best estimate of its actual size. Using this procedure, we have indeed found that the standard Stillinger cluster definition in general significantly overestimates the size of the critical cluster.[22]

IV. CONCLUSION

We have presented a powerful new method to analyze the kinetics of activated processes in simulation using the concept of MFPT. It provides us with a strikingly simple way to obtain and analyze all the quantities of major interest purely from the kinetics of the process. We could demonstrate the power of our method both analytically for a simple CNT barrier and for real MD simulation of LJ argon. One of the advantages of the method is that it can provide accurate kinetic measurements of the relevant quantities that can then be contrasted to pure equilibrium estimates. It can also be used to test the validity of different reaction coordinates. In addition, it provides a first-sight information about whether activation and relaxation/growth occur at similar time scales.

This MFPT method can be used with any reaction coordinate and it can also be generalized to more than one reaction coordinate. In addition, it can be applied to a myriad of different situations such as evaluating the rate and locating the transition state in a chemical reaction, the crystallization of liquids,[23] proteins,[24] or nanoclusters,[25] and in general any process where a barrier has to be overcome.

TABLE II. Comparison of the values of the nucleation τ_J and the critical size n^* for the same systems as in Fig. 4, but using two different Stillinger radii r_S.

Volume V	$(16\ \mathrm{nm})^3$	$(16\ \mathrm{nm})^3$	$(18\ \mathrm{nm})^3$	$(18\ \mathrm{nm})^3$
Stillinger radius r_S (σ_{LJ})	1.8	1.2	1.8	1.2
τ_J	11.1	10.4	157	154
n^*	13.2	10.3	14.1	11.7

ACKNOWLEDGMENTS

This work has been partially funded by German Academic Exchange Service (DAAD) and the Spanish DGCyT through the "Acciones Integradas Hispano-Alemanas" program.

APPENDIX: EVALUATION OF THE MFPT BY STEEPEST DESCENT

In this appendix we provide more details on the derivation of Eq. (8) from the MFPT Eq. (2) by the standard approximation of steepest descent.

In the case of a relatively high activation barrier, i.e., $\beta U(x^*) \gg 1$, the first exponential in the integrand of the MFPT

$$\tau(b) = \int_{x_0}^{b} \frac{1}{D_0} dy \exp[\beta U(y)] \int_{a}^{y} dz \exp[-\beta U(z)] \qquad \text{(A1)}$$

will be sharply peaked around the maximum of the free energy $U(y)$. Therefore, the overwhelming contribution to the integral of the MFPT would arise from a narrow size range in the vicinity of the transition state x^*. We can then expand the free energy $U(y)$ around the maximum x^*,

$$U(y) \simeq U(x^*) - \frac{1}{2}|U''(x^*)|(y-x^*)^2, \qquad \text{(A2)}$$

and replace D_0, which could in general depend on the reaction coordinate, by its value at the transition state $D_0(x^*)$, and the upper integration limit in the second integral also by x^*. The resulting expression is

$$\tau(b) = \frac{1}{D_0(x^*)} \int_{a}^{x^*} dz \exp[-\beta U(z)] \times \int_{x_0}^{b} dy \exp[\beta U(x^*)] \exp\left[-\frac{1}{2}\beta |U''(x^*)|(y-x^*)^2\right].$$

By changing the integration variable from y to $(y-x^*)$ and using the sharpness of the exponential to replace the

lower integration limit $(x_0 - x^*)$ by $-\infty$, the Gaussian integral in the previous equation can be reexpressed in terms of the error function $\mathrm{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{-x^2} dx$ as

$$\tau(b) = \frac{1}{D_0(x^*)} \exp[\beta U(x^*)] \times \int_a^{x^*} dz \exp[-\beta U(z)] \frac{1}{\sqrt{|U''(x^*)|/2\pi kT}} \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\sqrt{\frac{|U''(x^*)|}{2kT}}(b - x^*)\right)\right). \tag{A3}$$

The MFPT at the transition state is just

$$\tau(b = x^*) = \frac{1}{D_0(x^*)} \exp[\beta U(x^*)] \times \int_a^{x^*} dz \exp[-\beta U(z)] \frac{1}{\sqrt{|U''(x^*)|/2\pi kT}} \frac{1}{2} \equiv \frac{1}{2J}, \tag{A4}$$

where in the last step we have used Eq. (3). Using this result, we can finally simplify Eq. (A3) into

$$\tau(b) = \frac{\tau_J}{2}(1 + \mathrm{erf}((b - x^*)c)),$$

where $\tau_J = 1/J$, and $c = \sqrt{|U''(x^*)|/2kT}$.

[1] D. Reguera, J. M. Rubí, and J. M. G. Vilar, J. Phys. Chem. B **109**, 21502 (2005).
[2] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[3] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).
[4] J. Frenkel, *Kinetic Theory of Liquids* (Clarendon, Oxford, 1946), Chap. VII.
[5] K. F. Kelton, Solid State Phys. **45**, 75 (1991).
[6] J. H. ter Horst and D. Kashchiev, J. Chem. Phys. **119**, 2241 (2003).
[7] J. H. ter Horst and P. J. Jansens, Surf. Sci. **574**, 77 (2005).
[8] J. H. ter Horst and D. Kashchiev, J. Chem. Phys. **123**, 114507 (2005).
[9] L. S. Bartell and D. T. Wu, J. Chem. Phys. **125**, 194503 (2006).
[10] J. Wedekind, D. Reguera, and R. Strey, Poster at the Spring Meeting of the German Physical Society (DPG), Condensed Matter Division, Dresden, Germany, March 2006 (unpublished) (http://www.dpg-tagungen.de/archive/2006/dresden/index_en.html).
[11] D. Reguera and J. Wedekind; J. Wedekind, D. Reguera, and R. Strey, Talks Nos. 375 and 376 at the 80th ACS Surface and Science Symposium, Boulder, Colorado, June 2006 (unpublished) (www.colloid.org), where the first author of Ref. 9 was present and attended both talks.
[12] J. Wedekind, Ph.D. thesis, University of Cologne, 2006.
[13] J. Wedekind, D. Reguera, and R. Strey, J. Chem. Phys. **125**, 214505 (2006).
[14] R. McGraw, J. Phys. Chem. B **105**, 11838 (2001).
[15] P. G. Debenedetti, *Metastable Liquids* (Princeton University Press, Princeton, NJ, 1997).
[16] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1989).
[17] An empirical fit with similar results can also be done with a hyperbolic tangent function.
[18] P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, J. Chem. Phys. **104**, 9932 (1996); **110**, 1591 (1999); S. Auer and D. Frenkel, Nature (London) **413**, 711 (2001).
[19] C. Dellago, P. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[20] K. Yasuoka and M. Matsumoto, J. Chem. Phys. **109**, 8451 (1998).
[21] L. S. Bartell and G. Turner, J. Phys. Chem. B **108**, 19742 (2004).
[22] J. Wedekind, D. Reguera, and R. Strey (unpublished).
[23] I. Saika-Voivod, P. H. Poole, and R. K. Bowles, J. Chem. Phys. **124**, 224709 (2006).
[24] S. R. McGuffee and A. H. Elcock, J. Am. Chem. Soc. **128**, 12098 (2006).
[25] Y. G. Chushak and L. S. Bartell, J. Phys. Chem. B **105**, 11605 (2001).