Department Of Chemistry
University of Cambridge

Development of Boxed Molecular Dynamics for Efficient Simulations of Structural Transitions

Certificate of Postgraduate Studies Report

Boris Fačkovec
King's College

Supervisor:

David Wales

June 2013

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

Signature: ______________________________

Date: ______________________________

Acknowledgments

Scientific Support

David Wales, Joanne Carr

Financial Support

Bakala Foundation

Family

Lucia, Janka, Mária, Lenka

Abstract

This work presents a new method for calculating rate constants for configurational transitions described in terms of a master equation. The method is based on constraining molecular dynamics simulations to boxes in configuration space, and is also known as “boxed molecular dynamics”. Rate constants can be easily calculated even for systems deviating from an exponential distribution of the first passage times, as a result of the presence of internal barriers and roughness of the dividing surfaces. The theoretical justification of the method is based on the concept of mean first passage times. One of the assumptions of the reactive flux formulation is omitted; regression of the population evolution is used instead of calculation of the rate constant at a single point, so the distribution of first passage times is not required to be strictly exponential. The efficiency and correctness of the new method, boxed molecular dynamics in the first passage time formulation of the rate constants (FPT-BXD), is demonstrated for toy models. Preliminary results of simulations for a cluster of Lennard-Jones discs using FPT-BXD are discussed.

Glossary of Abbreviations

AXD	accelerated molecular dynamics
BXD	boxed molecular dynamics
DNEB	doubly nudged elastic band (method)
DPS	discrete path sampling
FPT	first passage time
FPT-BXD	boxed molecular dynamics in the first passage time formulation of the rate constants
HA	harmonic approximation
IDDT	intramolecular dynamics diffusion theory
LJ	Lennard-Jones
$\mathrm{LJ}_7^{\mathrm{2D}}$	cluster of seven Lennard-Jones disks
MD	molecular dynamics
ME	master equation
MFPT	mean first passage time
MSM	Markov state model
NGT	new graph transform
PES	potential energy surface
RMSD	root mean square displacement
RRKM	Rice, Ramsperger, Kassel and Marcus (theory)
SODE	system of ordinary differential equations
TPS	transition path sampling
TST	transition state theory

Glossary of Symbols

A	label of a box / species
$\mathbf{A}$	transition matrix
A_{ij}	an element of a transition matrix
a	dimensionless concentration of species A
B	label of a box / species
b	dimensionless concentration of species B
D	distance between two points in configuration space
E_A	activation energy
$F_{\mathrm{A}\to\partial\mathrm{AB}}$	normalised reaction flux from box A to dividing surface ∂AB
$\mathscr{F}_{\mathrm{A}\to\partial\mathrm{AB}}$	reaction flux from box A to dividing surface ∂AB
H	step function for box definition in configuration space
$\mathscr{H}$	Hamiltonian of the system
$\mathscr{H}_0$	Hamiltonian at the dividing surface
h	Planck's constant
$K_{\mathrm{A}\to\mathrm{B}}$	equilibrium constant
$k^{TST}_{\mathrm{A}\to\mathrm{B}}$	transition state theory rate constant
$k_{\mathrm{i}\to\mathrm{j}}$	rate constant of transition from box i to box j
$\mathrm{MFPT}_{\mathrm{A}\to\mathrm{B}}$	mean first passage time of a transition from A in equilibrium to B in equilibrium
$\mathrm{MFPT}_{\mathrm{A}\to\partial\mathrm{AB}}$	mean first passage time of a transition from A in equilibrium to dividing surface ∂AB
P	normalised population

$\mathscr{P}$	population
$\mathbf{p}$	coordinate vector
$\mathbf{q}$	momentum vector
$\mathbf{S}$	phase space
$\mathrm{d}s$	element of the dividing surface
$\mathscr{T}$	kinetic energy of the system
t	time
t_{i}^{fp}	first passage time for hitting box i starting from a point in a different box
u	velocity of a particle
$\mathscr{V}$	potential energy of the system
$V^{\dagger}$	minimum potential energy at the dividing surface
$v_{\mathrm{A}\to\mathrm{B}}$	reaction rate
W	total volume
Z	canonical partition function
$Z^{\ddagger}$	canonical sum of states at the dividing surface
β	$1/k_BT$
Γ_i	phase volume (partition function) of a box i
ϵ	well depth in the Lennard-Jones potential
$\varkappa$	reaction rate coefficient
κ	transmission coefficient
ρ	probability density at a point in phase space
σ	collision radius in the Lennard-Jones potential
$\tau^{\mathrm{A}\to\mathrm{B}}$	time of evolution of one trajectory starting from a phase state in set A ending in set B
φ	dynamics / phase flow
Ω	density of states

Contents

1 Introduction **1**

1.1 Master Equations 2

1.2 Rate Constants in Chemistry 4

1.3 Rate Constant Calculation 5

1.4 Simulation of Rare Events 8

1.5 Boxed Molecular Dynamics 10

2 Theory **13**

2.1 Definitions of Useful Quantities 13

2.2 Limitations of the Reactive Flux Approach 15

2.3 Irreversible Perturbation 17

2.4 A Reversible Reaction 19

2.5 Mean First Passage Time from Boxed Molecular Dynamics 20

2.6 Potential Sources of Error 22

3 Simulations **25**

3.1 Toy Models 25

3.1.1 Model Properties and Simulation Methods 25

3.1.2 Comparison of the TST and MFPT Approaches 28

3.1.3 Uneven Boxes 29

3.1.4 Higher Dimensions 30

3.2 Simulation of a Cluster of Lennard-Jones Disks 31

3.2.1 Characterisation of the System 31

3.2.2 Implementation 32

3.2.3 Preliminary Results 34

4 Conclusions and Future Work **35**

References **39**

List of Figures

1.1 Implementation of boxed molecular dynamics. 10

2.1 Comparison of species definitions. 14
2.2 Effect of surface roughness on the rate constant. 16
2.3 Comparison of systems with and without internal barriers. 17
2.4 Trajectories ending at the dividing surface. 18
2.5 Performance of TST and MFPT rate constants. 19
2.6 Reactive trajectories for a reversible transition. 20
2.7 Multi-box correlation 24

3.1 Exponential kinetics of the toy models. 26
3.2 Comparison of TST, harmonic TST, ME, BXD and FPT-BXD rate constants. 28
3.3 FPT-BXD results for a rough dividing surface. 29
3.4 FPT-BXD performance for uneven boxes. 30
3.5 Distribution of FPT for a box with many minima. 31
3.6 Disconnectivity graph of $\mathrm{LJ}_7^{\mathrm{2D}}$ 32
3.7 Partitioning of configuration space of $\mathrm{LJ}_7^{\mathrm{2D}}$ into boxes. 33
3.8 Energy profile of the configuration transition of $\mathrm{LJ}_7^{\mathrm{2D}}$. 34

Chapter 1

Introduction

Life is a non-equilibrium phenomenon. Biological processes result from complex networks of chemical reactions, diffusion and configurational transitions of molecules or supermolecular complexes. We are generally interested in **how** and **how fast** a particular process occurs. The question of **how** the process occurs stands for a qualitative information about the pathway and intermediates, which if modified, cause the nature and the rate of the process to change significantly. The question of **how fast** the process occurs concerns the quantitative description of the dynamics.

An example of an interesting biological process is protein folding, which is nature's solution to an NP-hard[1] (in some simplified formulations NP-complete[2;3]) complex non-linear optimisation problem. However, even simplified computer simulations on time scales of seconds using classical molecular dynamics (MD) would take thousands of years with modern computers. Simulation methods for more efficient simulations of dynamics of molecular systems have to be developed to make studying complex molecular systems feasible. In the last few decades, we have seen significant developments in methodology for simulating configurational transitions of molecular systems, ranging from small Lennard-Jones clusters to large biomolecules. Most of the methods are based on the reactive flux approach and transition state theory (TST) developed 80 years ago for chemical reactions, application of which to soft matter with low barriers results in systematic errors.

In the present work, the classical dynamics determined by the potential energy surface (PES)[4] is studied numerically. Species are defined as regions on the PES and the rates are defined based on the length of trajectories in the phase space. This chapter starts from a general formulation of deterministic dynamics of finite-dimensional systems and follows the approximations and the development leading to the previous formulation of boxed molecular dynamics (BXD).[5]

1.1 Master Equations

A dynamical system[6] is a tuple $\{\mathbf{S}, \varphi\}$, where $\mathbf{S}$ is the phase space and φ is a mapping $\varphi : \mathbf{S} \times \mathbb{R} \to \mathbf{S}$ satisfying two conditions:

$$\begin{aligned} \varphi(\mathbf{x}, 0) &= \mathbf{x} & \forall \mathbf{x} \in \mathbf{S} \\ \varphi(\varphi(\mathbf{x}, t), s) &= \varphi(\mathbf{x}, t+s) & \forall t, s \in \mathbb{R}, \forall \mathbf{x} \in \mathbf{S} \ . \end{aligned} \tag{1.1}$$

The physical meaning of the real number t in equation (1.1) is the evolution time between states $\mathbf{x}$ and $\varphi(\mathbf{x}, t)$. The first passage time (FPT) can be defined for each point $\mathbf{x}$ in $\mathbf{S}$ and a set of points $\mathrm{C} \subset \mathbf{S}$ as the minimum positive value of time t_{C}^{fp} satisfying the condition

$$\varphi(\mathbf{x}, t_{\mathrm{C}}^{fp}) \in \mathrm{C} \ . \tag{1.2}$$

The ultimate goal of studies in dynamics is to find an approximation of φ that is accurate and easy to evaluate. In the case of classical molecular systems, the phase space $\mathbf{S}$ is a product of an N-dimensional momentum and an N-dimensional configuration space. The dynamics φ (called also the phase flow) are given by a Hamiltonian vector field determined by the PES, generating an autonomous system of $2N$ ordinary differential equations (SODE):

$$\begin{aligned} \dot{q}_i &= \frac{p_i}{m_i} \\ \dot{p}_i &= -\frac{\partial \mathscr{H}(\mathbf{q}, \mathbf{p})}{\partial q_i} \ , \end{aligned} \tag{1.3}$$

where q_i and p_i are the i^{th} components of an N-dimensional spatial coordinate vector $\mathbf{p}$ and an N-dimensional momentum vector $\mathbf{p}$, respectively. m_i is the mass of the i^{th} particle. The Hamiltonian $\mathscr{H}$ generally consists of a non-linear function of coordinates $\mathbf{q}$ (potential energy $\mathscr{V}(\mathbf{q})$) and the kinetic energy $\mathscr{T}(\mathbf{p}) = \sum_i p_i^2/2m_i$. Evaluation of phase flow in constant time is possible only exceptionally for such a system. The dynamics are usually simulated by a numerical integration of the SODE (1.3), which scales linearly with the length and the number of the simulated trajectories.

A common approach to simplify Equation (1.3) is to discretise the phase space into boxes. In chemistry, this approach is widely used. Configuration space boxes correspond to species (molecules, ions, radicals etc.) and kinetic equations describe the evolution of the populations in the boxes. If only unimolecular reactions occur, which is the case for structural transitions, the system (1.3) of kinetic equations reduces to a linear homogeneous SODE:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \ , \tag{1.4}$$

where each component of vector $\mathbf{x}$, x_i, is a population of the i^{th} box and $\mathbf{A}$ is the transition matrix specific to the system and the discretisation. Equation (1.4) has an analytical solution

$$\mathbf{x}(t) = e^{\mathbf{A}t} \, \mathbf{x}(0) \ . \tag{1.5}$$

The right-hand side of equation (1.5) can be evaluated accurately and quickly for reasonably large systems (up to millions of boxes).

In the context of conformational transitions, equation (1.4) is called the master equation.[7;8] If the space is divided into n boxes, information about the dynamical behaviour of the system is reduced to $n^2 - n$ elements of $\mathbf{A}$, also called the rate constants. The answer to the question of **how** the process occurs is given by the boxes undergoing a significant population change during the process. In chemistry, the set of discrete paths through the boxes is called the reaction mechanism. The question of **how fast** the process occurs is answered by the value of the overall rate constant, which has to be evaluated by a simulation of the system or using other approximations. Models for studying the evolution of the populations in the configuration space boxes are special cases of Markov state models (MSM) and have been recently successfully used for studying biomolecules.[9;10]

Discretisation of the phase space is a good approximation if the transition from one box to another follows exponential kinetics. In phase space, such behaviour implies an exponential distribution of the time lengths of the reactive trajectories. This distribution usually applies if the species are separated by high energy barriers. However, the presence of high energy barriers dividing the species is neither a necessary nor a sufficient condition. Another important requirement of the approximation is the decorrelation of input and output trajectories. The behaviour of the species must be independent on the reaction in which it was produced.[11] Every discretisation should be checked for the applicability of the MSM,[12] for example by comparing the internal equilibration time[13] with the characteristic time of the transition. More detailed theory and applications of MSM's to biomolecules can be found in methodology papers[14–16] and recent reviews.[8;17]

1.2 Rate Constants in Chemistry

A classical approach to describe unimolecular elementary reactions by rate equations and to calculate the rate constant given the evolution proceeds as follows. Let us consider a simple elementary irreversible reaction

$$\mathrm{A} \rightarrow \mathrm{B} \ . \tag{1.6}$$

The dimensionless concentration $a(t) = [\mathrm{A}](\mathrm{t})/[\mathrm{A}](0)$ follows power law kinetics

$$v_{\mathrm{A}\rightarrow\mathrm{B}}(t) = \frac{\mathrm{d}b(t)}{\mathrm{d}t} = -\frac{\mathrm{d}a(t)}{\mathrm{d}t} = k_{\mathrm{A}\rightarrow\mathrm{B}}\ a(t) \ , \tag{1.7}$$

with boundary conditions

$$a(0) = 1, \ \lim_{t\rightarrow\infty} a(t) = 0 \ , \tag{1.8}$$

where $v_{\mathrm{A}\rightarrow\mathrm{B}}(t)$ is the reaction rate at time t and $k_{\mathrm{A}\rightarrow\mathrm{B}}$ is the rate constant. Solving the differential equation (1.7) leads to

$$a(t) = e^{-k_{\mathrm{A}\rightarrow\mathrm{B}}t} \tag{1.9}$$

and

$$v_{\mathrm{A}\rightarrow\mathrm{B}}(t) = k_{\mathrm{A}\rightarrow\mathrm{B}}e^{-k_{\mathrm{A}\rightarrow\mathrm{B}}t} \ . \tag{1.10}$$

The rate constant $k_{\mathrm{A}\rightarrow\mathrm{B}}$ can be calculated from the evolution of species A as:

$$k_{\mathrm{A}\rightarrow\mathrm{B}} = -\frac{\frac{\mathrm{d}}{\mathrm{d}t}a(t)}{a(t)} \ , \tag{1.11}$$

which is independent of time t in the case of exponential behaviour for $a(t)$. The rate constant can be calculated by fitting the evolution of $a(t)$ with (1.9) or by evaluating the reciprocal of the mean value of $v_{\mathrm{A}\rightarrow\mathrm{B}}(t)$:

$$k_{\mathrm{A}\rightarrow\mathrm{B}} = \frac{\displaystyle\int_0^\infty v_{\mathrm{A}\rightarrow\mathrm{B}}(t)\mathrm{d}t}{\displaystyle\int_0^\infty t\ v_{\mathrm{A}\rightarrow\mathrm{B}}(t)\mathrm{d}t} \ . \tag{1.12}$$

Let us now discuss the same system with a backward reaction,

$$\mathrm{A} \rightleftharpoons \mathrm{B} \ . \tag{1.13}$$

The concentration of species A evolves in time as

$$a(t) = \frac{k_{\mathrm{B\to A}}}{k_{\mathrm{A\to B}} + k_{\mathrm{B\to A}}} + \frac{k_{\mathrm{A\to B}}}{k_{\mathrm{A\to B}} + k_{\mathrm{B\to A}}} e^{-(k_{\mathrm{A\to B}} + k_{\mathrm{B\to A}})t} , \tag{1.14}$$

and the rate of reaction as

$$v_{\mathrm{A\to B}}(t) = k_{\mathrm{A\to B}} e^{-(k_{\mathrm{A\to B}} + k_{\mathrm{B\to A}})t} . \tag{1.15}$$

Knowing the equilibrium constant of the system

$$K_{\mathrm{AB}} = \frac{k_{\mathrm{A\to B}}}{k_{\mathrm{B\to A}}} , \tag{1.16}$$

we can calculate the rate constant with a method analogous to (1.11) as the solution of equations (1.16) and

$$\frac{k_{\mathrm{A\to B}}}{k_{\mathrm{A\to B}} + k_{\mathrm{B\to A}}} = -\frac{\frac{\mathrm{d}}{\mathrm{d}t} a(t)}{a(t)} . \tag{1.17}$$

By analogy with system (1.6), the rate constant can be calculated without solving a system of algebraic equations by fitting $a(t)$ with (1.14) or using equation (1.12).

1.3 Rate Constant Calculation

The scientific development of the theory of chemical dynamics dates back to the 19th century when van't Hoff studied the dependence of reaction rate on temperature[18] and Arrhenius introduced the concept of activation energy.[19] In the following 50 years, great advances were achieved. Farkas first used the concept of equilibrium flux to calculate reaction rates.[20] Eyring introduced the concept of an "activated complex",[21] the saddle point on the PES connecting the reactant and the product. He derived a formula for the "absolute" rate constant for a reaction of any order

$$k_{\mathrm{A\to B}}^{\mathrm{Eyring}}(T) = \kappa \frac{1}{\beta h} \frac{Z^{\ddagger}}{Z} e^{-\beta E_A} , \tag{1.18}$$

where κ is the transmission coefficient, an *ad hoc* parameter being generally about unity, and $Z^{\ddagger}$ and Z are the partition sums of the activated state and the reactant, respectively. $\beta = 1/(k_B T)$ where T is the thermodynamic temperature and k_B is the Boltzmann constant, h is Planck's constant and E_A is the activation energy. The reactant, activated complex and product are explicitly defined as single points on the PES.

Transition state theory[22–24] (TST) provides the fundamental basis for the reactive flux method used predominantly today. In TST, the rate constant is defined as the equilibrium flux through the dividing surface divided by the population inside the reactant box. In 1938, Wigner summarised[25] the assumptions of TST:

1. the adiabatic separation of the movements of the electrons and nuclei (the Born-Oppenheimer[26] approximation),
2. the motion of the nuclei can be described by classical mechanics,
3. all trajectories crossing the dividing surface are reactive (no recrossing of the dividing surface occurs).

Transition state theory is inherently a classical mechanical theory applicable for reactions in which a transition over a state with high energy is the determining step.[23] From Wigner's paper,[23] it can be inferred that he does not define species unambiguously by the dividing surface. The equilibrium rate constant can be calculated as an integral over the dividing surface in phase space satisfying the non-recrossing condition:

$$k_{\mathrm{A}\to\mathrm{B}}^{\mathrm{TST}}(T) = \frac{W^2}{Z}\int \frac{\mathrm{d}\mathscr{H}_0(\mathbf{q},\mathbf{p})/\mathrm{d}t}{|\nabla\mathscr{H}_0(\mathbf{q},\mathbf{p})|}\mathrm{d}s \ , \tag{1.19}$$

where $\mathscr{H}_0(\mathbf{q},\mathbf{p}) = 0$ defines the surface and $\mathrm{d}\mathscr{H}_0(\mathbf{q},\mathbf{p})/\mathrm{d}t$ is

$$\begin{aligned}\frac{\mathrm{d}\mathscr{H}_0(\mathbf{q},\mathbf{p})}{\mathrm{d}t} = &\sum_i \left(\frac{\partial\mathscr{H}_0(\mathbf{q},\mathbf{p})}{\partial q_i}\frac{\partial(\mathscr{H}(\mathbf{q},\mathbf{p})-\mathscr{H}_0(\mathbf{q},\mathbf{p}))}{\partial p_i} - \right.\\ &\left. -\frac{\partial\mathscr{H}_0(\mathbf{q},\mathbf{p})}{\partial p_i}\frac{\partial(\mathscr{H}(\mathbf{q},\mathbf{p})-\mathscr{H}_0(\mathbf{q},\mathbf{p}))}{\partial q_i}\right) \ .\end{aligned} \tag{1.20}$$

The "total volume" W in expression (1.19) is used to scale the rate constant which Wigner derived for a trimolecular reaction.

The classical TST rate constant in the microcanonical ensemble was developed by Rice,[27] Ramsperger,[28] Kassel[29] and Marcus[30] (RRKM). The microcanonical rate constant for transitions from A to B, can be written as

$$k_{\mathrm{A}\to\mathrm{B}}^{\mathrm{TST}}(E) = \frac{g(E)}{h\ \Omega_{\mathrm{A}}(E)} \ , \tag{1.21}$$

where E is the total energy, Ω_{A} is the density of states of box A, and $g(E)$ is defined as

$$g(E) = \int_{V^\ddagger}^{E}\Omega^\ddagger(E')\mathrm{dE}' \ , \tag{1.22}$$

where $\Omega^\ddagger(E')$ is the density of states at the dividing surface and $V^\ddagger$ is the minimum potential energy of the transition state ensemble. From the relationship between the microcanonical and canonical ensembles it follows that $k(T)$ is the Laplace transform of $k(E)$.

The third assumption was soon identified as the main cause of the divergence between the TST rate constants and the rate constants obtained from experiments. Chandler reformulated the rate constant[31] in the formalism of correlation functions[32] using Onsager's hypothesis.[33;34] The simulation method based on this formula is known as the "Bennett-Chandler" procedure[35] and is usually performed in two steps. First, the TST rate constant is calculated. Second, the TST rate constant is corrected by the transmission coefficient κ:

$$k^{\mathrm{BC}}_{\mathrm{A}\to\mathrm{B}} = \kappa \; k^{\mathrm{TST}}_{\mathrm{A}\to\mathrm{B}} \; . \tag{1.23}$$

κ is calculated from the probability of recrossing obtained from simulations of trajectories starting at the dividing surface.

Another approach to correct for recrossings, variational TST,[36;37] is based on the assumption that the optimum dividing surface is the one that minimises the recrossings. New insights were brought by studies identifying the transition state ensemble with the hypersurface in the configuration space with the probability to reach products (committor) equal to 0.5,[38] and studies of phase space using a normally hyperbolic invariant manifold[39;40] for construction of the transition state surface. Kramers studied motion of a Brown particle in a potential field[41] and derived analytical formulae for the high and low friction limits. His results, generalised by Grote and Hynes,[42] were later shown[43] to be equivalent to TST for parabolic barriers. More information on recent developments of TST can be found in topic reviews. [44–46]

Another approach to compute rate constants is the calculation of the mean first passage times (MFPT's). Instead of studying the equilibrium flux, actual trajectories and their evolution times are studied. Bunker and Hase studied the distribution of FPT's (in their terminology "gap times")[47] and showed that even chemical reactions do not follow strictly exponential distributions. In the microcanonical ensemble, there is a non-negligible ensemble of periodic trajectories that do not escape from their box. Behaviour deviating from the statistical RRKM description was studied by plotting histograms of FPT's by Hase and co-workers.[48] However, most of the development of the MFPT approach has been considered in configuration space, the MFPT being the solution to a partial differential equation derived from the Smoluchowski equation.[49] The reciprocal of MFPT (in the configuration space formulation) and the TST rate constant were shown to be equivalent for the high barriers and well-defined dividing surfaces.[49;50]

1.4 Simulation of Rare Events

A rare event can be loosely defined as a process that would take too much time to simulate by conventional methods. The existence of rare events arises from the fact that some systems show interesting behaviour on time scales much larger than the shortest vibrational time determined by the structure of their PES. Examples of such events are protein folding, conformational transitions of large biomolecules, ions passing through a membrane channel, chemical reactions and many others. General approaches to simulating rare events include: freezing uninvolved degrees of freedom,[35;51] using multiple time steps,[52] parallelisation,[53] parallel tempering,[54] biasing the potential,[55] and modification of the scaling behaviour with barrier height from exponential to polynomial by reformulating the sampling from an initial value problem to a boundary value problem.[56]

The unprecedented advancement in computational power over the last 30 years gave rise to new, more efficient methods for the simulation of rare events. While calculation of energy profiles along a selected coordinate is well understood and relatively reliable,[57] calculation of rate constants is usually based on TST, which provides the upper limit for the reaction rate. Here only the methods most relevant to this work are briefly explained. More details about the methods can be found in recent reviews[58;59] or Danielle Moroni's thesis.[60]

Perhaps the most advanced method for calculation of classical rate constants is transition path sampling (TPS).[61–63] By analogy with Metropolis Monte Carlo,[64] the sampling of transition paths can be efficient because only small steps are made from already known highly probable paths. The original formulation of random walks in the transition path space was extended to deterministic dynamics.[65] TPS not only provides highly accurate estimates of rate constants, but also the most probable transition path. The efficiency can be improved by defining more dividing surfaces, leading to a similar method, transition interface sampling.[60;66] TPS has also been recently improved to overcome barriers in path space more easily.[67] Nevertheless, the computational cost of simulating a sufficient number of transition paths limits its general usability.

Constraining the dynamics to a subset of the phase space can enforce simulation of the desired rare event. In the Blue Moon method,[68;69] the system is constrained in a hypersurface in the configuration space and the mean force perpendicular to this surface is calculated. Paci and Ciccotti used the method to calculate the transmission coefficient for vacancy migration in a Lennard-Jones crystal.[70] The formula for the free energy was later modified so that it contains only explicit variables[71] and was extended for the use of a general vectorial coordinate.[72] Other examples of methods based on constraining MD are accelerated dynamics[73] introduced by Shalashilin and co-workers, and boxed molecular dynamics,[5] which constrains the dynamics in a box in configuration space.

Many other accelerated molecular dynamics methods have been proposed. Hyperdynamics[74] fills boxes with a biasing potential and the times of processes on the resulting shallower potential are renormalised accordingly. Various biasing potentials have recently been used.[75;76] Temperature accelerated MD[77] calculates the rate constants at higher temperatures and extrapolates to low temperature assuming the Arrhenius equation.

In milestoning[78] short simulations between predefined hypersurfaces in configuration space (called milestones) are performed instead of a single long one. The dynamical behaviour is calculated from statistical properties of the short trajectories. The method gives accurate results if isocommittor surfaces (committor = 0.5) are used as the milestones.[79] Milestoning with boxes defined by Voronoi tesselation [80] can be generally applied without any knowledge of the best reaction coordinate. In Voronoi partitioning of the space, a set of box centres $\mathbf{x}_i$ uniquely defines the boxes. A point in space $\mathbf{x}$ belongs to box A if

$$D(\mathbf{x}, \mathbf{x}_A) < w(A, i)D(\mathbf{x}, \mathbf{x}_i), \quad \forall i \neq A \ , \tag{1.24}$$

where $D(\mathbf{x}, \mathbf{y})$ means the distance between points $\mathbf{x}$ and $\mathbf{y}$. The root mean square distance (RMSD) is usually used as the measure D. In classical Voronoi tesselation, $w(i, j)$ is equal to 1 for any pair of boxes i, j.

Significantly increased computational efficiency can be achieved with discrete path sampling[81–83] (DPS) for systems with a reasonably small numbers of low-lying minima. Transition states between the minima are found by geometry optimisation[84–86] and the rate constants between neighbouring minima are calculated by an appropriate method, commonly the TST approach. The harmonic approximation can be used to allow fast estimation of the TST rate constants. A framework for systematic improvement of the rate constants would be beneficial. Phenomenological rate constants between the reactant and product sets of states can be calculated using the new graph transform procedure[87] (NGT) in which species defined as the basins of attraction of local minima are gradually removed while the transition matrix is renormalised.

1.5 Boxed Molecular Dynamics

One of the recently proposed methods for simulating the thermodynamics and kinetics of rare events, boxed molecular dynamics[5] (BXD), combines the advantages of two older methods: intramolecular diffusion dynamics theory[88;89] (IDDT) and molecular dynamics accelerated by phase space constraints[73] (AXD). In IDDT, the configuration space is sliced along a reaction coordinate and the diffusion coefficient is calculated using short MD simulations. In AXD, the reactant configuration space is divided into two boxes. One box is placed close to the dividing surface representing the transition state ensemble and the other box represents the reactants. In BXD, this approach is generalised to more boxes placed along the reaction coordinate. The free energy profile is calculated from the flux ratios between the boxes. BXD aims to efficiently simulate the dynamics of the process using the master equation with the rates calculated from the flux values.

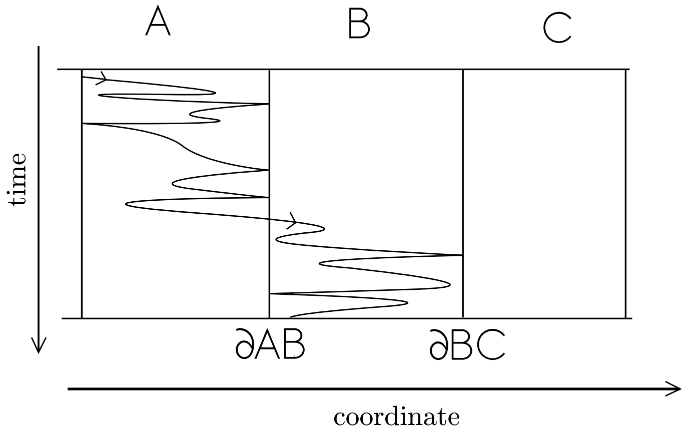

Figure 1.1: Implementation of boxed molecular dynamics according to Glowacki *et al.*[5] After a few (up to 30) inversion events, a hit from box A can be used for initialisation of the trajectory in box B. The method is not parallelised for the sake of easier initialisation of the trajectory in a box.

Implementation of BXD is straightforward. An independent MD simulation is run in each box. If the trajectory leaves the box, the positions are returned to the previous point of the simulation and a velocity inversion procedure is performed. The projections of velocities onto the reaction coordinate are inverted while the velocity components parallel to the dividing surface remain unchanged. The velocities of atoms not involved in the reaction coordinate are unchanged. This inversion procedure does not change the energy, momentum or angular momentum. The time of each inversion event is recorded. The BXD trajectory in box i can be considered as an ensemble of $n_{\mathrm{i}-1} + n_{\mathrm{i}+1}$ subtrajectories, $n_{\mathrm{i}-1}$ being inverted from (ending at) the boundary between boxes i and i-1 and $n_{\mathrm{i}+1}$ from the boundary between boxes i and i+1. The rate constant for the transition from box i to box i-1 is calculated from the simulation as the equilibrium flux:

$$k^{eq}_{\mathrm{i}\to\mathrm{i}-1} = \frac{n_{\mathrm{i}-1}}{\sum_{j=1}^{n_{\mathrm{i}-1}} \tau_j} \ , \tag{1.25}$$

where τ_j is the time length (evolution time) of the j^{th} trajectory inverted from the boundary between i and i-1. The free energy difference between boxes i and i-1 is calculated from simulations in both boxes as:

$$\Delta G_{\mathrm{i}-1\to\mathrm{i}} = -k_B T \ln\left(\frac{k^{eq}_{\mathrm{i}-1\to\mathrm{i}}}{k^{eq}_{\mathrm{i}\to\mathrm{i}-1}}\right) \ . \tag{1.26}$$

Global dynamics can be described using the master equation with a tridiagonal transition matrix $\mathbf{A}$ with non-zero elements

$$\begin{aligned} A_{i,i-1} &= k^{eq}_{\mathrm{i}-1\to\mathrm{i}} \ , \\ A_{i,i} &= -k^{eq}_{\mathrm{i}\to\mathrm{i}-1} - k^{eq}_{\mathrm{i}\to\mathrm{i}+1} \ , \\ A_{i,i+1} &= k^{eq}_{\mathrm{i}+1\to\mathrm{i}} \ . \end{aligned} \tag{1.27}$$

The authors explicitly state that the detailed balance condition must be satisfied and the dynamics must be ergodic for BXD to give correct results. The BXD approach also assumes that the inversion procedure does not disturb the equilibrium distribution in the boxes. An implicit assumption of BXD is that the TST rate constants can be used for transitions between neighbouring boxes. However, the validity of these assumptions has not been tested separately. The correctness of BXD as a whole is demonstrated by consistency of the results with "brute force" MD and milestoning. Free energy profiles have been shown to be robust with respect to box selection with fast convergence of the results with the number of subtrajectories $n_{\mathrm{i}-1} + n_{\mathrm{i}+1}$ sampled in each box i.

BXD was used to study conformational changes of small (10-13 amino acid) peptides.[5] The gain in computational efficiency was clearly demonstrated. BXD has the potential to significantly decrease the computational cost of simulating the dynamics of rare events. Firstly, slicing the reaction coordinate into boxes can significantly reduce the barrier height. Secondly, the independence of the simulations in the boxes makes parallelisation of the method trivial and formally correct. Another advantage of the method is its natural relationship with the master equation. BXD can be used for specialised applications, such as rationalisation of the power law dynamics of loop formation in a small peptide.[90]

A later paper[91] by the same authors, which focused more on dynamics, suggested corrections for fast dynamical motion. The distributions of FPT's differ from those obtained using milestoning and are not exponential. The artificial increase in the number of short trajectories leads to overestimation of the calculated rate constants. The authors suggest setting an evolution time threshold and excluding the trajectories below the threshold. A more systematic correction of the rate constants would significantly improve the method. Other important issues also have to be resolved. The assumptions of BXD should be tested separately using simple models, and a method of error estimation should be developed. Voronoi tesselation can be used to define the boxes instead of a reaction coordinate.[91] The method currently uses the Langevin equation, so it depends on an unphysical friction constant. Using deterministic MD would systematically improve the description of the dynamics. The present work attempts to benchmark and further develop BXD.

Chapter 2

Theory

2.1 Definitions of Useful Quantities

In the energy landscapes view, a “species” is a useful yet artificial concept discretising the configuration space. A proper definition of species is essential for accurately describing any process on the landscape. For example, in TPS[92] studies of a cluster of Lennard-Jones discs, small spheres in RMSD space surrounding the minima were used. A convenient definition used in DPS assigns regions in configuration space to the basins of attraction of particular local minima.[4] Dividing surfaces then roughly agree with the maxima on the transition pathways. However, assignment of a structure to the corresponding minimum can be computationally expensive and the number of boxes also corresponds to the number of minima, which grows roughly exponentially with the number of particles. Another method of partitioning the space is Voronoi tesselation in which phase points are assigned to one of the boxes using a selected measure. As well as partitioning along a collective coordinate, the Voronoi method seems to be the most convenient for master equation modelling since it describes all states of the system, so the sum of the populations is conserved in any process, and in principle it allows an arbitrary number of boxes. The dividing surface is unambiguously defined and the transmission coefficient is unity (see figure 2.1).

In the master equation, a transition from any state A to a different state B is assumed to be a Poisson process. Any memory of previous processes is completely lost. Therefore, a standard distribution of states inside a particular box must be defined. Any property of species A can be calculated as the mean property of all states in box A using this distribution. A natural, and perhaps the most convenient choice, is the equilibrium distribution. A transition from box A to box B can then be described by an ensemble of trajectories with starting points evenly (in the

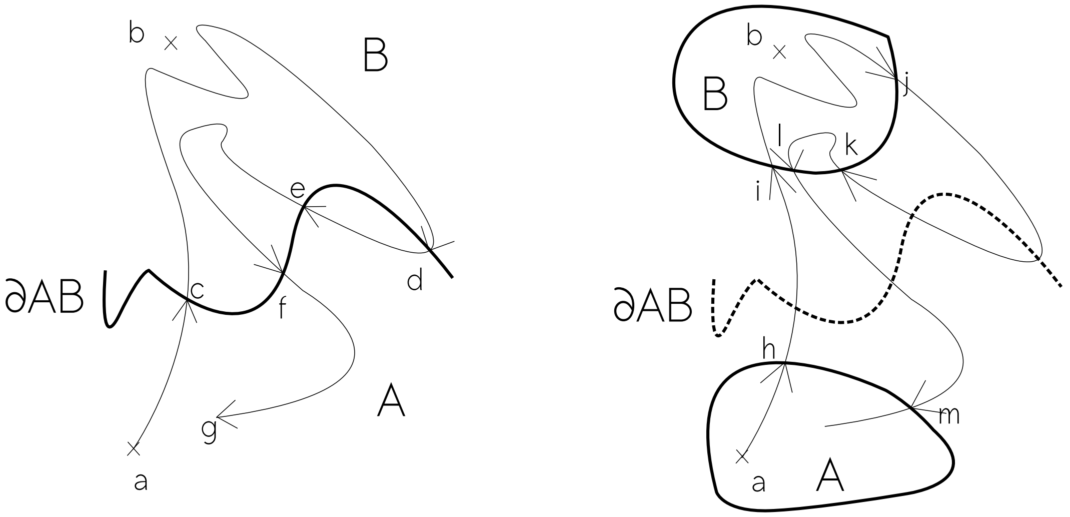

Figure 2.1: Comparison of the species definition used in BXD (left) and TPS (right). Thick full line defines the species, dashed line represents the dividing surface and the thin line represents a trajectory obtained from a simulation. Points a and b are minima on the PES. Points c to m divide the trajectories into subtrajectories. **Left:** All trajectories crossing the dividing surface (a-c, c-d, d-e and e-f) except for f-g are reactive since at their ends the particle becomes a different species. The transmission coefficient is unity for a simulation that ends at the dividing surface (a-f). **Right:** h-i and l-m are reactive trajectories and j-k is a non-reactive trajectory. The total population of states at times the particle is outside the boxes (h-i, j-k, l-m) is lower than at times the particle is inside either of the boxes (a-h, i-j, k-l).

microcanonical ensemble) distributed in A and endpoints evenly distributed in B.

Now let us define some useful quantities used in the following discussions in microcanonical ensemble. The box phase volume, Γ_{A}, is

$$\Gamma_{\mathrm{A}} = \int_{\mathbf{q}\in A} 1 \ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q} = \int H(\mathbf{q}, A) \ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q} \ , \tag{2.1}$$

where $H(\mathbf{q}, A)$ is one if $\mathbf{q}$ lies in A and zero otherwise. The equilibrium probability of being in box A, P_{A}, is given simply by the ratio of its volume to the total volume of all the boxes

$$P_{\mathrm{A}} = \frac{\Gamma_{\mathrm{A}}}{\sum_{\mathrm{i}} \Gamma_{\mathrm{i}}} \ . \tag{2.2}$$

The non-equilibrium population of states in box A at time t is given by

$$\mathscr{P}_{\mathrm{A}}(t) = \int_{\mathbf{q}\in A} \rho(\mathbf{p}, \mathbf{q}, t) \ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q} = \int \rho(\mathbf{p}, \mathbf{q}, t) \ H(\mathbf{q}, A) \ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q} \ , \tag{2.3}$$

where $\rho_{eq}(\mathbf{p}, \mathbf{q}, t)$ is the probability of state $(\mathbf{p}, \mathbf{q})$ at time t. The probability of being in box A at time t, which is the central quantity for master equation modelling, is

$$P_{\mathrm{A}}(t) = \frac{\mathscr{P}_{\mathrm{A}}(t)}{\sum_{\mathrm{i}} \mathscr{P}_{\mathrm{i}}(t)} \ . \tag{2.4}$$

For each phase point $(\mathbf{p}, \mathbf{q})$ in A we can define the first passage time (FPT) $t^{fp}_{\mathrm{A}\to\partial\mathrm{AB}}$ as the time it takes for the trajectory starting from $(\mathbf{p}, \mathbf{q})$ to reach any phase point in the boundary ∂AB. The flux through the dividing surface between A and B, ∂AB, at time t can be defined by the population of states as follows: let us consider a system in which configuration space is divided into two boxes A and B only. Let all the trajectories leaving B through the dividing surface at time t_0 be reflected back to box B. The rate of change of population of states in B is given by the flux through the boundary surface ∂AB

$$\mathscr{F}_{\mathrm{A}\to\partial\mathrm{AB}}(t_0) = \frac{\partial}{\partial t} \mathscr{P}_{\mathrm{B}}(t)|_{t=t_0} \ . \tag{2.5}$$

The normalised flux can be defined in a similar way as:

$$F_{\mathrm{A}\to\partial\mathrm{AB}}(t_0) = \frac{\partial}{\partial t} P_{\mathrm{B}}(t)|_{t=t_0} = \frac{\mathscr{F}_{\mathrm{A}\to\partial\mathrm{AB}}(t_0)}{\mathscr{P}_{\mathrm{A}}(t_0)} \ . \tag{2.6}$$

This flux between the boxes in equilibrium is used in TST for the definition of the rate constants. The rate coefficient $\varkappa_{\mathrm{A}\to\partial\mathrm{AB}}$ can be defined at time t_0 as

$$\varkappa_{\mathrm{A}\to\partial\mathrm{AB}}(t_0) = \frac{\mathscr{F}_{\mathrm{A}\to\partial\mathrm{AB}}(t_0)}{\mathscr{P}_{\mathrm{A}}(t_0)} = \frac{F_{\mathrm{A}\to\partial\mathrm{AB}}(t_0)}{P_{\mathrm{A}}(t_0)} \ . \tag{2.7}$$

Master equation modelling assumes this quantity to be independent of time. If we assume that the states in B are in equilibrium with the others lying on the same trajectory at all times, the flux $F_{\mathrm{A}\to\partial\mathrm{AB}}(t_0)$ is formally equivalent to the reaction rate $v_{\mathrm{A}\to\mathrm{B}}$ [see equation (1.10)] and $\varkappa_{\mathrm{A}\to\partial\mathrm{AB}}(t_0)$ is the rate constant.

2.2 Limitations of the Reactive Flux Approach

Using the TST rate constants to describe the dynamics in terms of the master equation faces two main problems. First, a strong dependence on the dividing surface cannot be corrected by a transmission coefficient if the boxes touch, since then $\kappa \equiv 1$. Second, the evolutions (1.9) and (1.14) may depend on internal barriers which are disregarded in the TST approach.

Roughness of the Dividing Surface

In TST, the rate constant, which should be an average over the whole box defining the reactants, strongly depends on the dividing surface, which represents only a small subset of the ensemble. Miller argues[93] that using the characteristic function χ for the reaction, the TST rate constant does not depend on the dividing surface. However, he does not explicitly discuss the definition of species. From the possibility of many dividing surfaces it follows that species are not defined as touching boxes in configuration or phase space, as there must be a sufficient gap allowing decorrelation.

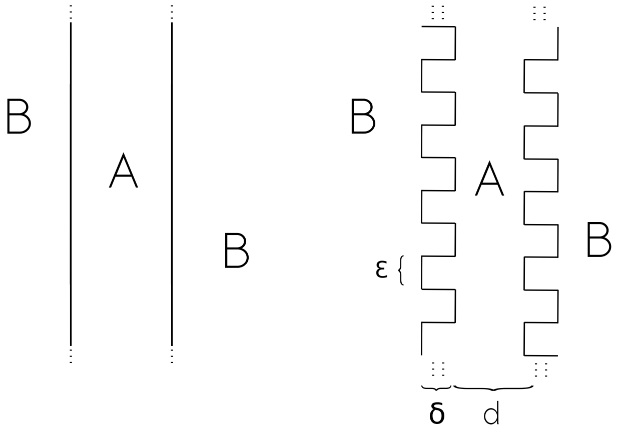

Figure 2.2: **Left:** box A with a smooth dividing surface between A and B. **Right:** box A with a rough dividing surface between A and B. $k^{TST}_{\mathrm{A}\to\mathrm{B}}$ will be higher, even if $\delta \ll d$.

To illustrate the effect of a rough dividing surface on the dynamics, let us discuss two systems. First, let us divide a plane into boxes A and B by two parallel lines of infinite potential (Figure 2.2 left). Without loss of generality, let the potential energy be everywhere constant in A, $\mathscr{V} \equiv 0$. A point particle moving in A with velocity u will bounce from either of the walls approximately once in time $t = d/u$, where d is the width of box A. Now let us increase the roughness (and therefore the length) of the dividing surface by lamellae of length $\delta \ll d$ and width $\varepsilon \ll \delta$. We will observe a series of many (roughly $2\delta/\varepsilon$) bounces separated in time by $t' = \varepsilon/u$ once in $t = d/u$. The number of hits per time unit and therefore the apparent reactive flux will be much higher, but the real evolution of the system should not change since the change in the definition of species was negligible.

The Effect of Internal Barriers

In region A we can define any number of internal dividing surfaces. Trajectories starting from states far from the boundary of A must cross many of these inner surfaces, and these crossings and these crossings could correspond to a much slower process than the crossing of ∂AB. Region A can contain internal barriers (energetic or entropic, such as spacial bottlenecks or mazes, see figure 2.3). Hence the mean passing time through the region can be higher than the time needed for crossing the final barrier. The equilibrium rate constant does not need to include these effects. However, in a description of real dynamics, this neglect is equivalent to the assumption that the flux through the dividing surface is the rate-determining event of the whole transition.

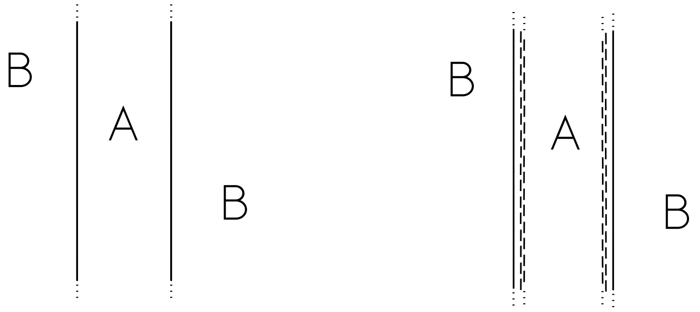

Figure 2.3: **Left:** box A without an internal barrier, where crossing the outer dividing surface may be the rate limiting step. **Right:** box A with an internal barrier (maze) with a low probability of crossing.

2.3 Irreversible Perturbation

For the reasons explained in the previous section, $\varkappa_{\mathrm{A}\rightarrow\partial\mathrm{AB}}$ is rarely independent of time in real systems. Let us discuss the irreversible reaction (1.6) generalised to phase space. Let the configuration space be divided into two boxes, A and B. Let all the microstates of the system be in equilibrium. At time $t = 0$ we irreversibly disturb the equilibrium, so that the population of states in box B is zero:

$$\begin{aligned} \rho(\mathbf{p},\mathbf{q},0) &\equiv 0 \quad \forall(\mathbf{p},\mathbf{q}) \notin A \ , \\ \rho(\mathbf{p},\mathbf{q},0) &\equiv 1 \quad \forall(\mathbf{p},\mathbf{q}) \in A \ . \end{aligned} \tag{2.8}$$

From time $t = 0$ we let the system evolve in a way such that no trajectory can leave box B (Figure 2.4).

Let us assume that box A has some internal barriers and that the dividing surface between A and B is rough. The $P(t)$ decay will be steeper at low time values because of surface roughness and slower at high time values because of internal barriers. The rate constant describing the dynamics during the whole process can be obtained by fitting the evolution to an exponential function using least squares, leading to the equation:

$$\frac{\partial}{\partial k_{\mathrm{A}\rightarrow\partial\mathrm{B}}} \int_0^\infty \left(P(t) - e^{-k_{\mathrm{A}\rightarrow\partial\mathrm{B}} t}\right)^2 \,\mathrm{d}t = 0 \ , \tag{2.9}$$

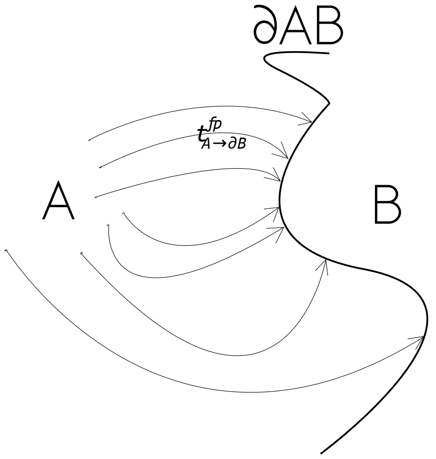

Figure 2.4: Illustration of the reactive trajectories. Each trajectory starts at a phase state in A and ends at the boundary of box B. The distribution of trajectories entering B (normalised flux $F_{\mathrm{A}\rightarrow\partial\mathrm{AB}}$) is equal to the distribution of first passage times.

which can be solved numerically. The reactive flux method in this case leads to:

$$k^{TST}_{\mathrm{A}\rightarrow\partial\mathrm{B}} = \frac{F(0)}{P(0)} \ , \tag{2.10}$$

which is generally different from the expression (2.9) for the rate constant the best describes the process.

Although the reciprocal of the MFPT is not generally equal to k in equation (2.9), it can be a good approximation of the rate constant. First, the MFPT involves information about $P(t)$ for all times t. Second, the MFPT often identified with the "waiting time" is explicitly required by some methods for dynamical simulation, including the kinetic Monte Carlo[94–96] and DPS approach. The MFPT is the ensemble average of the time it takes for the equilibrium distribution to pass through the dividing surface:

$$\mathrm{MFPT}_{\mathrm{A}\rightarrow\partial\mathrm{AB}} = \frac{\displaystyle\int t^{fp}_{\mathrm{A}\rightarrow\partial\mathrm{AB}}(\mathbf{p},\mathbf{q})\ H(\mathbf{p},\mathbf{q},\mathrm{A})\ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}}{\displaystyle\int H(\mathbf{p},\mathbf{q},\mathrm{A})\ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}} \ , \tag{2.11}$$

where $H(\mathbf{p},\mathbf{q},\mathrm{A})$ is 1 if $(\mathbf{p},\mathbf{q})$ is in A and 0 otherwise. Let us now compare how rate constants defined as the equilibrium flux and the reciprocal value of the MFPT fit a non-exponential $P(t)$. A good test function is

$$P(t) = c_1 e^{-c_2 t} + (1 - c_1)e^{-t} \ . \tag{2.12}$$

Low values of parameter c_1 represent small deviations from a strictly exponential distribution. High values of the parameter c_2 represent deviations resulting from surface roughness, while low values model the effect of internal barriers. From figure 2.5 it is obvious that the rate constant definition based on equilibrium flux fails to describe the long-time dynamics even for small deviations. The MFPT rate constant can be used for very rough surfaces and very high internal barriers (values of c_2 very small or very large, respectively) provided the proportion of the affected phase space (c_1) is small.

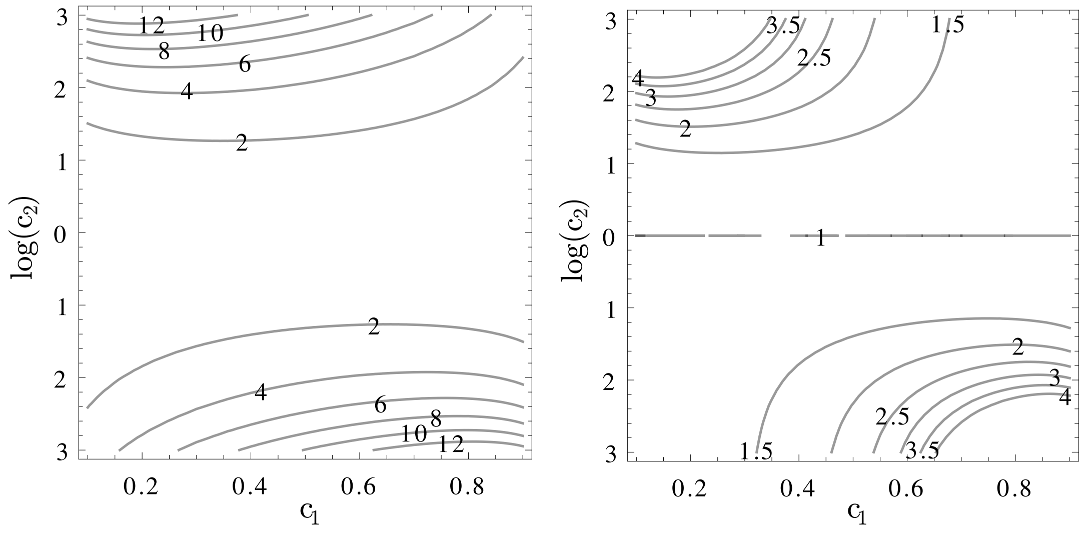

Figure 2.5: Quality of the description of long-time dynamics using TST (left) and MFPT (right) rate constants as a function of c_1 and $\log(c_2)$ from (2.12) given as the ratio of the TST (MFPT) rate constants to the rate constant obtained analytically as the solution to (2.9). Isosurface contours represent values of k 1.5 to 4 times higher for the MFPT and 2 to 12 times higher for TST. The MFPT rate constant generally lies much closer to the best value.

2.4 A Reversible Reaction

Now let us consider a reversible reaction $A \rightleftharpoons B$. At time $t = 0$ we irreversibly disturb the equilibrium, so that the population of states is distributed according to equation (2.8). Fitting the evolution (1.14) leads to rate constants that are not generally consistent with the exit MFPT rate constants derived in the previous section. It is obvious from figure 2.3 that the change of an infinitesimally small part of box A does not affect the equilibrium between box A and B but it can affect the $\mathrm{MFPT}_{\mathrm{A}\rightarrow\partial\mathrm{AB}}$. Therefore, the rate constants calculated from the exit MFPT will not have the property (1.16).

The inconsistency originates in the neglect of the penetration time in the calculation of $\mathrm{MFPT}_{\mathrm{A}\rightarrow\mathrm{B}}$. The whole transition from box A in equilibrium to B in

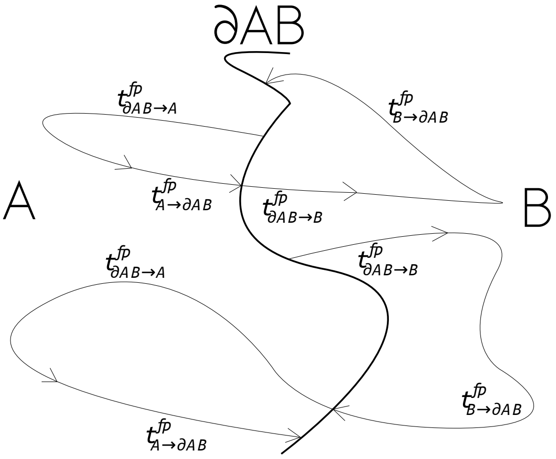
∂AB
$t^{fp}_{\partial AB\to A}$
$t^{fp}_{B\to\partial AB}$
A
$t^{fp}_{A\to\partial AB}$
$t^{fp}_{\partial AB\to B}$
B
$t^{fp}_{\partial AB\to A}$
$t^{fp}_{\partial AB\to B}$
$t^{fp}_{B\to\partial AB}$
$t^{fp}_{A\to\partial AB}$

Figure 2.6: Illustration of reactive trajectories. Each trajectory starting at ∂AB going through box A returning to ∂AB can be divided into penetration $\partial\text{AB} \to \text{A}$ and exit $\text{A} \to \partial\text{AB}$ trajectories.

equilibrium can be divided into two consecutive processes. First, trajectories must leave box A by reaching the boundary ∂AB. Then, they propagate from ∂AB to B, such that their end points are distributed according to equilibrium probability in B. All trajectories $\partial\text{AB} \to \text{A} \to \partial\text{AB}$ can be therefore divided into exit and penetration trajectories. If the phase points in A are in equilibrium with other phase points lying on the same trajectory in A throughout the whole process, the ratio of leaving to penetration times is given by the ratio of the corresponding equilibrium fluxes:

$$\frac{F_{\mathrm{A}\to\mathrm{B}}(t)}{P_{\mathrm{A}}(t)}\,\mathrm{MFPT}_{\mathrm{A}\to\partial\mathrm{AB}} = \frac{F_{\mathrm{B}\to\mathrm{A}}(t)}{P_{\mathrm{B}}(t)}\,\mathrm{MFPT}_{\partial\mathrm{AB}\to\mathrm{A}}\ . \tag{2.13}$$

The rate constant calculated as

$$k_{\mathrm{A}\to\mathrm{B}} = \frac{1}{\mathrm{MFPT}_{\mathrm{A}\to\mathrm{B}}} = \frac{f_{\mathrm{AB}} + f_{\mathrm{BA}}}{f_{\mathrm{AB}}\left(\mathrm{MFPT}_{\partial\mathrm{AB}\to\mathrm{A}\to\partial\mathrm{AB}} + \mathrm{MFPT}_{\partial\mathrm{AB}\to\mathrm{B}\to\partial\mathrm{AB}}\right)}\ , \tag{2.14}$$

where $f_{\mathrm{ij}} = F_{\mathrm{i}\to\mathrm{j}}(t)/P_{\mathrm{i}}(t)$, gives the correct equilibrium distribution. The reactive flux approach can give the correct equilibrium properties but the MFPT's have to be used for calculating the rate constants.

2.5 Mean First Passage Time from Boxed Molecular Dynamics

As discussed in the previous section, the MFPT does not generally need to be the optimal fit of $P(t)$. Provided that a sufficient number of subtrajectories is sampled, the rate constant of exiting the box can be directly fitted to $P(t)$ obtained from the simulation as:

$$P(t) = 1 - \frac{\displaystyle\int_0^t n(\tau)\ \mathrm{d}\tau}{\displaystyle\int_0^\infty n(\tau)\ \mathrm{d}\tau}\ , \tag{2.15}$$

where $n(\tau)$ is the number of sampled subtrajectories shorter than τ. However, the MFPT can be more convenient for the reasons discussed above.

For a sufficiently long trajectory, the ensemble average is equal to the time average in an ergodic system:

$$\langle X \rangle = \frac{\displaystyle\int X(\mathbf{p},\mathbf{q})\ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}}{\displaystyle\int 1\ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}} = \frac{\displaystyle\int X(\mathbf{p}(t),\mathbf{q}(t))\ \mathrm{d}t}{\displaystyle\int 1\ \mathrm{d}t}\ , \tag{2.16}$$

where X is a quantity defined for each phase state, in our case the first passage time. The denominator in the latter fraction is simply the time length (evolution time) of the trajectory, τ. Let us consider two boxes A and B in the configuration space. The average over box A is

$$\langle X \rangle_\mathrm{A} = \frac{\displaystyle\int X(\mathbf{p},\mathbf{q})\ H(\mathbf{q},\mathrm{A})\ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}}{\displaystyle\int H(\mathbf{q},\mathrm{A})\ \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}} = \frac{\displaystyle\int X(\mathbf{p}(t),\mathbf{q}(t))\ H(\mathbf{q}(t),\mathrm{A})\ \mathrm{d}t}{\displaystyle\int H(\mathbf{q}(t),\mathrm{A})\ \mathrm{d}t}\ , \tag{2.17}$$

where $H(\mathbf{q},\mathrm{A})$ is one if $\mathbf{q}$ belongs to A and zero otherwise. The denominator is the time length of the part of the trajectory lying in A.

Calculation of the MFPT proceeds as follows. Propagation of a trajectory can be started (time $t = 0$) from any phase state in A. The simulation is stopped immediately after it hits ∂AB, so it has time length $\tau^{\mathrm{A}\to\partial\mathrm{AB}}$. The first passage time is defined for each point as

$$t^{fp}_{\mathrm{A}\to\partial\mathrm{AB}}(\mathbf{p}(t),\mathbf{q}(t)) = \tau^{\mathrm{A}\to\partial\mathrm{AB}} - t\ . \tag{2.18}$$

For all phase states that lie on the trajectory and in box A, we can calculate the MFPT:

$$\mathrm{MFPT}_{\mathrm{A}\to\partial\mathrm{AB}} = \frac{\displaystyle\int_0^{\tau^{\mathrm{A}\to\partial\mathrm{AB}}} (\tau^{\mathrm{A}\to\partial\mathrm{AB}} - t)\ H(\mathbf{q}(t),\mathrm{A})\ \mathrm{d}t}{\displaystyle\int_0^{\tau^{\mathrm{A}\to\partial\mathrm{AB}}} H(\mathbf{q}(t),\mathrm{A})\ \mathrm{d}t}\ . \tag{2.19}$$

Now let us consider a configuration space divided into two boxes A and B, so that every trajectory escaping from A ends in B. A and B touch, so the propagation of the trajectory is stopped at the same time as it leaves box A. The MFPT calculated from one such trajectory is

$$\langle t^{fp}_{\mathrm{A}\to\partial\mathrm{AB}}\rangle_{\mathrm{1traj}} = \frac{\displaystyle\int_0^{\tau^{\mathrm{A}\to\partial\mathrm{AB}}} (\tau^{\mathrm{A}\to\partial\mathrm{AB}} - t)\ \mathrm{d}t}{\displaystyle\int_0^{\tau^{\mathrm{A}\to\partial\mathrm{AB}}} 1\ \mathrm{d}t} = \frac{\tau^{\mathrm{A}\to\partial\mathrm{AB}}}{2}\ . \tag{2.20}$$

For a sample of n trajectories, the MFPT is a time-weighted average of first passage times

$$\mathrm{MFPT}_{\mathrm{A}\to\partial\mathrm{AB}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \tau_{\mathrm{i}}^{\mathrm{A}\to\partial\mathrm{AB}} \langle \mathrm{t}^{fp}_{\mathrm{A}\to\partial\mathrm{AB}}\rangle_{\mathrm{i}}}{\sum_{\mathrm{i}=0}^{\mathrm{n}} \tau_{\mathrm{i}}^{\mathrm{A}\to\partial\mathrm{AB}}}\ . \tag{2.21}$$

Combining equations (2.20) and (2.21) gives

$$\mathrm{MFPT}_{\mathrm{A}\to\partial\mathrm{AB}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\tau_{\mathrm{i}}^{\mathrm{A}\to\partial\mathrm{AB}})^2}{2\sum_{\mathrm{i}=1}^{\mathrm{n}} \tau_{\mathrm{i}}^{\mathrm{A}\to\partial\mathrm{AB}}}\ . \tag{2.22}$$

The equilibrium (TST) and average (MFPT) rate constants become equal for exactly exponential distributions of FPT's. Consistently, formulae (1.25) and (2.22) then become identical in the limit of an infinite number of sampled trajectories. For an infinite number of sampled trajectories, we can write equation (2.22) as

$$\mathrm{MFPT}_{\mathrm{A}\to\partial\mathrm{AB}} = \frac{k\displaystyle\int_0^\infty \tau^2\ e^{-k\tau}}{2k\displaystyle\int_0^\infty \tau\ e^{-k\tau}} = \frac{\frac{2k}{k^3}}{\frac{2k}{k^2}} = \frac{1}{k}\ . \tag{2.23}$$

Equation (1.25) implies that the MFPT is $1/k$:

$$\mathrm{MFPT}^{\mathrm{TST}}_{\mathrm{A}\to\partial\mathrm{AB}} = \frac{k\displaystyle\int_0^\infty \tau\ e^{-k\tau}}{2\ k\displaystyle\int_0^\infty e^{-k\tau}} = \frac{\frac{k}{k^2}}{1} = \frac{1}{k}\ . \tag{2.24}$$

2.6 Potential Sources of Error

Let us collate the assumptions of classical theories for calculation of rate constants using the definitions from section 2.1:

1. the motion of electrons and nuclei can be separated (accordingly to the Born-Oppenheimer approximation),

2. the motion of nuclei can be described by classical mechanics,
3. the distribution of phase points within a selected box is in equilibrium,
4. the trajectories entering a box are statistically independent of the trajectories leaving the box,
5. a) the population in each box i is in equilibrium with every other box j,

   b) the distribution of FPT's is strictly exponential,

   c) the distribution of FPT's is a non-increasing continuous function.

Both the method of reactive flux and the MFPT based method presented in this work require assumptions 1-4 to be valid. Assumption 5.b) is a weaker form of assumption 5.a) and is required by equilibrium flux methods. An even weaker assumption, 5.c) is required by the MFPT approach used here. An FPT-BXD simulation protocol that correctly reproduces classical dynamics cannot be in conflict with assumptions 3, 4 and 5 and the sampling must be ergodic.

The inversion procedure used by Glowacki *et al.*[5] can sometimes give incorrect equilibrium distributions. If the potential energy gradient points towards the box boundary, BXD trajectories moving almost tangentially to the dividing surface cannot leave the neighbourhood of the boundary because they are reflected from the boundary at the same small angle. Such trajectories will unphysically increase the probability of states close to the boundary and therefore increase the calculated flux. To allow sampling of trajectories that penetrate deeper into the box, the velocity directions of all particles should be randomised. A new inversion method is tested in the present work.

As discussed in section 2.4, BXD simulations in both neighbouring boxes A and B are necessary to calculate the rate constant corresponding to transitions from A to B. The sum of the MFPT's has to be divided into boxes according to the equilibrium fluxes (2.14). However, without internal barriers the uncorrected and corrected rate constants can be almost equal. Since the ratio of equilibrium fluxes calculated from simulations can be inaccurate, the correction can be omitted if the recrossing error dominates. Such a simplification must be checked by plotting the distribution of lag times.

If the configuration space is divided into two boxes representing the species only, the rate constant calculated by fitting $P(t)$ obtained from equation (2.15) and scaled by equilibrium fluxes using equation (2.15) gives in principle the exact classical rate constant for the system. However, the efficiency gain of BXD is based on the possibility of introducing new boxes. Division of the space into more boxes results in less assigned volume per box and therefore smaller distances between the box boundaries. An insufficient distance between two neighbouring boxes can cause the incoming trajectories to be correlated with trajectories leaving the box. The decorrelation is one of the requirements of MSM's (assumption 4 above) and can be mathematically formulated as

$$k_{\mathrm{i}\rightarrow\mathrm{j}\rightarrow\mathrm{k}} = k_{\mathrm{l}\rightarrow\mathrm{j}\rightarrow\mathrm{k}} \ , \tag{2.25}$$

where i, k, l are boxes neighbouring box j and $k_{\mathrm{i}\rightarrow\mathrm{j}\rightarrow\mathrm{k}}$ is the rate constant $k_{\mathrm{j}\rightarrow\mathrm{k}}$ under the constraint that trajectories can enter box j only from box i. The condition can be written as

$$\mathrm{MFPT}_{\partial\mathrm{ij}\rightarrow\partial\mathrm{jk}} = \mathrm{MFPT}_{\partial\mathrm{lj}\rightarrow\partial\mathrm{jk}} \ . \tag{2.26}$$

Here $\mathrm{MFPT}_{\partial\mathrm{ij}\rightarrow\partial\mathrm{jk}}$ is the mean first passage time of those trajectories exiting box j through dividing surface ∂jk, which entered box j through dividing surface ∂ij. Such data can easily be obtained from FTP-BXD simulations by recording the endpoints and evolution times of all subtrajectories. This procedure was not possible with the old formula in which recrossing and the equality of weight of long and short trajectories in the summation would cause recrossing to artificially decrease the $\mathrm{MFPT}_{\partial\mathrm{kj}\rightarrow\partial\mathrm{jk}}$ compared to other $\mathrm{MFPT}_{\partial\mathrm{ij}\rightarrow\partial\mathrm{jk}}$. Analysis of the correlation in the simulation can indicate that the box is too small ($\mathrm{MFPT}_{\partial\mathrm{ij}\rightarrow\partial\mathrm{jk}} < \mathrm{MFPT}_{\partial\mathrm{kj}\rightarrow\partial\mathrm{jk}}$) or that the box contains a high internal barrier ($\mathrm{MFPT}_{\partial\mathrm{kj}\rightarrow\partial\mathrm{jk}} < \mathrm{MFPT}_{\partial\mathrm{ij}\rightarrow\partial\mathrm{jk}}$).

Even if $\mathrm{MFPT}_{\partial\mathrm{kj}\rightarrow\partial\mathrm{jk}} = \mathrm{MFPT}_{\partial\mathrm{ij}\rightarrow\partial\mathrm{jk}}$, the Markovian assumption can be broken by an inappropriate configuration of boxes. A high internal barrier could spread throughout more boxes (see figure 2.7). However, such cases will probably not be commonplace.

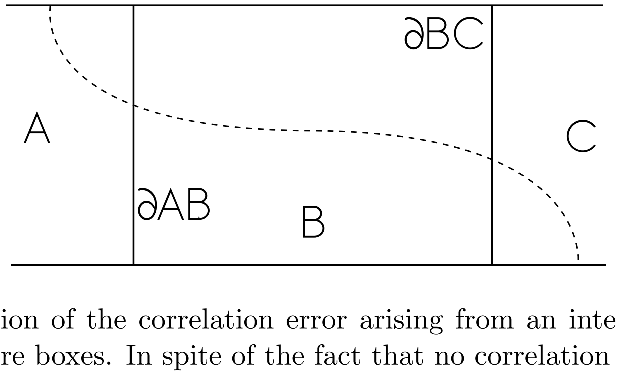

Figure 2.7: Illustration of the correlation error arising from an internal barrier (dashed line) spread over more boxes. In spite of the fact that no correlation can be implied from a BXD simulation in box B, the transition from box A to box C will be much smaller than calculated from a BXD simulation ($\mathrm{MFPT}_{\mathrm{A}\rightarrow\mathrm{C}} > \mathrm{MFPT}_{\mathrm{A}\rightarrow\mathrm{B}} + \mathrm{MFPT}_{\mathrm{B}\rightarrow\mathrm{C}}$), so assumption 4 is violated. The method proposed in this section cannot identify such a pathological case.

Chapter 3

Simulations

Simulations of toy models were carried out to prove the correctness and to test the efficiency of FPT-BXD for the calculation of rate constants for master equation (ME) modelling. Variations of a simple two-dimensional periodic potential with a single minimum per box are used in section 3.1. A many-dimensional generalisation of this potential is used in section 3.1.4 to demonstrate transferability of the results to realistic systems with multiple minima per box. FPT-BXD calculation of rate constants for a cluster of seven Lennard-Jones discs has not yet been successful, for reasons that are understood. The current state of the simulation protocol optimisation is described in section 3.2.

3.1 Toy Models

3.1.1 Model Properties and Simulation Methods

The dynamics of the transitions between two square boxes in a simple two-dimensional periodic potential of the form

$$\mathscr{V}(x, y) = \cos(2\pi x) + \cos(2\pi y) \tag{3.1}$$

were examined. Such a potential has many periodic orbits and the motion in x is completely independent of the motion in the y direction and *vice versa*. Therefore, a randomisation potential was introduced by adding randomisation lines. If such a line is crossed, the direction of the velocity vector is rotated by an angle uniformly distributed between 0 and 2π. The line can be physically interpreted as an infinitely thin wall of alternating charges, which changes the direction of a passing charged particle. Box A is defined as the region $-1 < x < 0$ and $0 < y < 1$. Box B is defined by $0 < x < 1$ and $0 < y < 1$. The boundary conditions can be viewed as a simulation on the surface of a torus. A particle exiting box A or B through the line $y = 1$ simply enters the same box from the line $y = 0$, while its velocity remains unchanged. The boxes have two boundaries, $x = 1$ (which is in an analogous way connected to $x = -1$) and $x = 0$. The randomisation lines were folded to 5 overlapping circles with radii of 0.2 located at [0.3,0.3], [0.3,0.7], [0.5,0.5], [0.7,0.3] and [0.7,0.7] in box B and at the equivalent positions in box A.

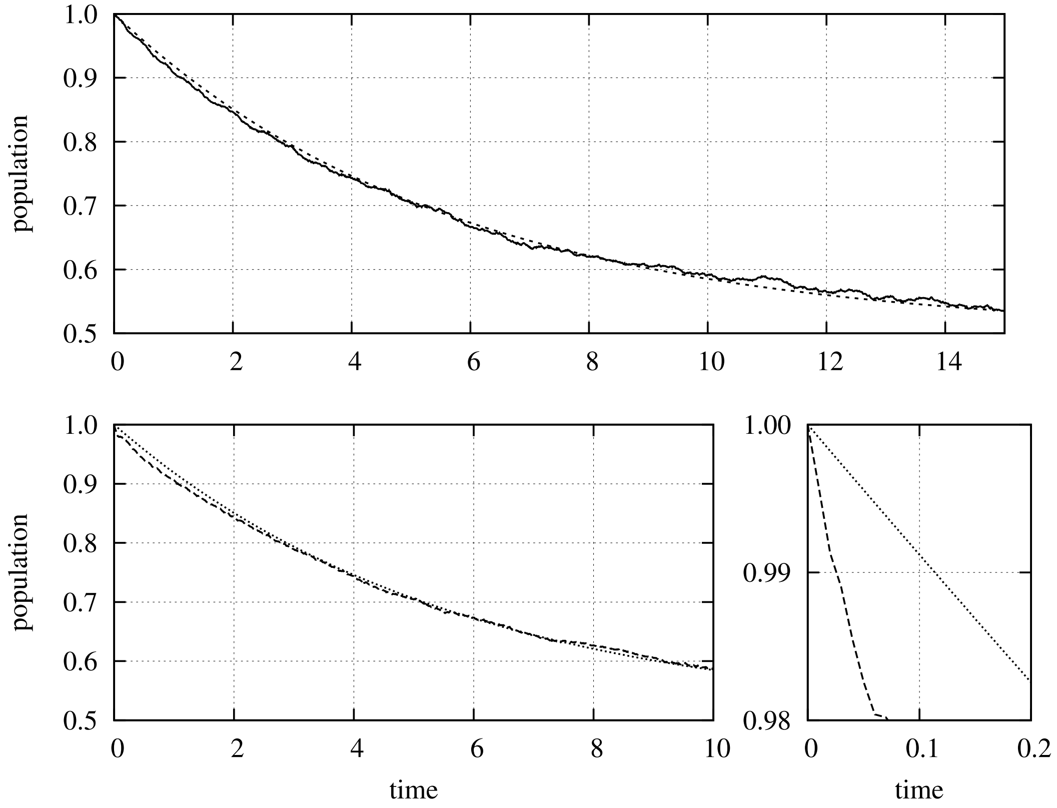

Figure 3.1: Evolutions of the populations obtained from the ME simulations for the smooth (top) and the rough (bottom left) dividing surfaces. The dashed (top) and dotted (bottom) lines were obtained by fitting the ME evolutions to an exponential function. A steep decay corresponding to recrossing can be found for the rough dividing surface at small time values (bottom right). Nevertheless, ME dynamics of the used toy models are exponential even for high energies (E=0.5). The models are suitable for description in terms of the master equation.

Classical microcanonical simulations using the velocity Verlet integrator[97] with time step dτ=0.01 were used both for BXD and ME simulations. The ME simulation here stands for a direct simulation of relaxation of the population of 10^3 independent particles in box A from the state in which population in A equals 10^3 and in B equals 0. In both BXD and ME simulations, the particles were first equilibrated in a system comprising both boxes for 10^4 steps. In ME simulations of the A $\rightarrow$ B transition process, all particles in B were deleted and the system was propagated for another $5 \cdot 10^4$ steps. The concentration defined as the number of particles in box A divided by the number of particles in both boxes was recorded in each step. The evolution was fitted to an exponential function (1.14). As seen in figure 3.1, the ME dynamics for the system are exponential even for the rough dividing surfaces [$x = 0.05 \sin(20\pi y)$ and $x = 1 + 0.05 \sin(20\pi y)$]. BXD simulations were equilibrated in the same way. Then, the simulations were constrained in the desired boxes for another $5 \cdot 10^4$ steps. A particle hitting the dividing surface, for example in case of the smooth surface $x = -1$ or $x = 0$ for box A and $x = 0$ or $x = 1$ for box B, was returned to its previous position and the velocity direction was randomised, requiring only that the sign of its x component pointed into the box. The statistical error of each calculated rate constant was estimated as the standard deviation of values obtained from 15 independent simulations. The simulation protocols were implemented and vectorised in Octave 3.2.[98]

The microcanonical TST rate constant was calculated using the RRKM formula (1.21). Densities of states of both the dividing surface and the box were calculated numerically and fitted to a polynomial expansion. Each configuration space box was discretised into a square grid of 10^6 equally distant points, and the density of states was calculated for each one for energies between -2 and 2. The integral density of states $g(E)$ was then fitted with a linear function. The function

$$g(E) = 3.653 + 4.321E \tag{3.2}$$

describes the exact $g(E)$ very well for energies between 0 and 0.5. Both the smooth and the rough surfaces were discretised into grids of points with the distances between the neighbouring points set to 10^{-4}. The smooth surface can be well described with the function

$$g^{\ddagger}(E) = 0.177E \tag{3.3}$$

for energy values between 0 and 0.5. The integral density of states for the rough surface can be approximated with a polynomial fit as

$$g^{\ddagger}(E) = 0.008 + 0.359E + 0.013E^2 \ . \tag{3.4}$$

The rate constants were calculated as

$$k^{TST}_{\mathrm{A}\to\mathrm{B}}(E) = 2 \times 2\pi \times \frac{g^{\ddagger}(E)}{\frac{\mathrm{d}}{\mathrm{d}E} g(E)} \ , \tag{3.5}$$

where the first factor of 2 follows from presence of two dividing surfaces between the boxes. The numerical result for the smooth surface agrees well with the RRKM rate constant in the harmonic approximation:

$$k^{HA}_{\mathrm{A}\to\mathrm{B}}(E) = \Xi \frac{\nu^{\eta}_{\mathrm{A}}}{\nu^{\ddagger(\eta-1)}} \left(\frac{E - V^{\ddagger}}{E - V_{\mathrm{A}}} \right)^{\eta-1} , \tag{3.6}$$

which can be calculated for the toy model used in this work as

$$k^{HA}_{\mathrm{A}\to\mathrm{B}}(E) = \frac{2E}{E+2} . \tag{3.7}$$

Here $\Xi = 1$ is a degeneracy factor for the reaction, $\eta = 2$ is the number of degrees of freedom, $\nu_{\mathrm{A}} = 1$ is the harmonic vibrational frequency in the minimum of box A with energy $V_{\mathrm{A}} = -2$, and $\nu^{\ddagger} = 1$ at the harmonic frequency in the lowest energy point of the transition state ensemble, with energy $V^{\ddagger} = 0$. The factor 2 in (3.7) follows from the presence of two dividing surfaces ($x = 0$ and $x = -1$).

3.1.2 Comparison of the TST and MFPT Approaches

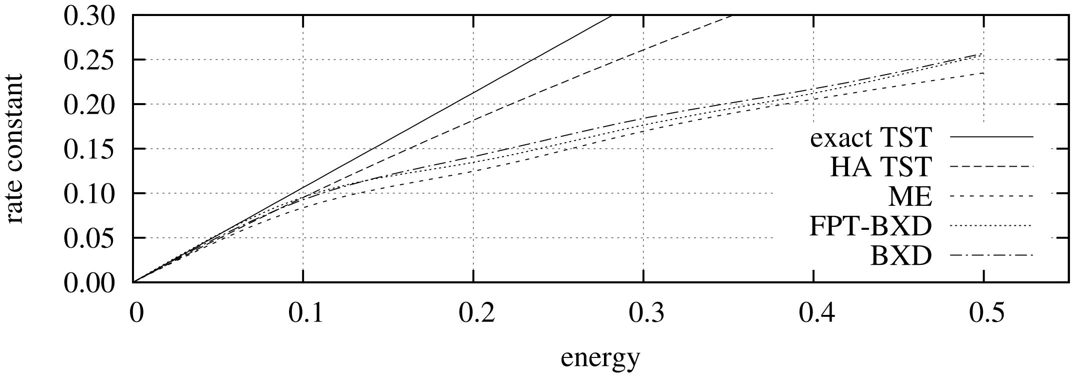

Figure 3.2: Comparison of ME (the lowest dashed curve) rate constants for the smooth dividing surface with the rate constants from BXD simulations using the reactive flux (BXD) and the MFPT (FPT-BXD) formulation and with the exact RRKM (exact TST) and RRKM in the harmonic approximation (HA TST).

The results for ME and BXD simulations were calculated for energy values 0, 0.1, 0.2, 0.3, 0.4 and 0.5. Since the statistical errors are very small ($\Delta k_{\mathrm{A}\to\mathrm{B}} < 0.005$), the curves can be fitted with cubic splines (figure 3.2). The rate constants calculated using the old and the new formulations of BXD agree well with the reference ME data for the smooth dividing surface. The TST rate constants significantly differ from the reference ME rate constants at higher energies, where crossing the barrier is not the rate limiting process. The rate constants all converge to 0 as $E \to 0$ and they seem to be consistent for low energies, $\lim_{E\to 0} k^{i}_{\mathrm{A}\to\mathrm{B}}/k^{j}_{\mathrm{A}\to\mathrm{B}} = 1$ (figure 3.2). A dividing surface defined by the curve $x = 0.05\sin(20\pi y)$ (the total length of dividing surface ≈ 2.3 instead of 1 in the case of the smooth dividing surface) was used to demonstrate the large increase of the reactive flux rate constant with dividing surface roughness. Comparison of the rate constants calculated using the new and the old BXD formulations and the ME demonstrates that the reactive flux approach significantly overestimates the rate constant if the dividing surface is rough (figure 3.3). The old formulation of BXD gives higher rate constants than the MFPT formulation. The TST rate constant is non-zero for zero energy since the dividing surface contains points with $E < 0$.

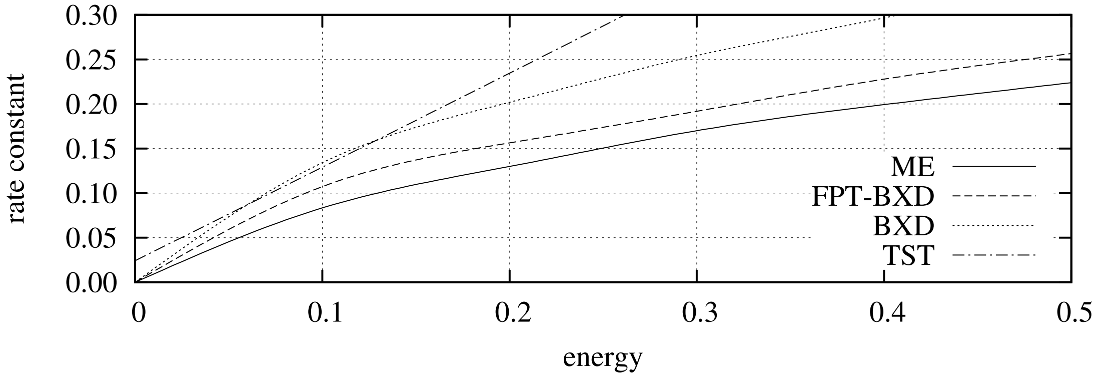

Figure 3.3: Dependence of the rate constants on energy for the rough dividing surface. The exact TST rate constant (dash-dot line) [see equation (3.4)] differs significantly from the reference ME (solid line) rate constant. The BXD rate constant based on the reactive flux formulation (dotted line) is generally higher than the FPT-BXD rate constant (dashed line). Since the statistical error is below 0.005, the data for E=0, 0.1, 0.2, 0.3, 0.4 and 0.5 were fitted with cubic splines.

3.1.3 Uneven Boxes

An asymmetric term was added to study the effect of uneven equilibrium populations

$$\mathscr{V}(x,y) = \cos(2\pi x) + \cos(2\pi y) + 0.2\sin(\pi x) \ . \tag{3.8}$$

The equilibrium constant and the reference rate constants were obtained by fitting the ME evolution. Exit MFPT's for boxes were obtained from BXD simulations. The rate constants calculated from exit MFPT's using equation (2.14) agree with the reference ME rate constants (figure 3.4).

The projection of the density of states into configuration space (the probability distribution of coordinates) was studied for two constraining methods: bouncing from the wall according to Glowacki *et al.*[5] and randomisation of the velocity direction. As predicted in section 2.6, the inversion procedure by Glowacki *et al.*[5]

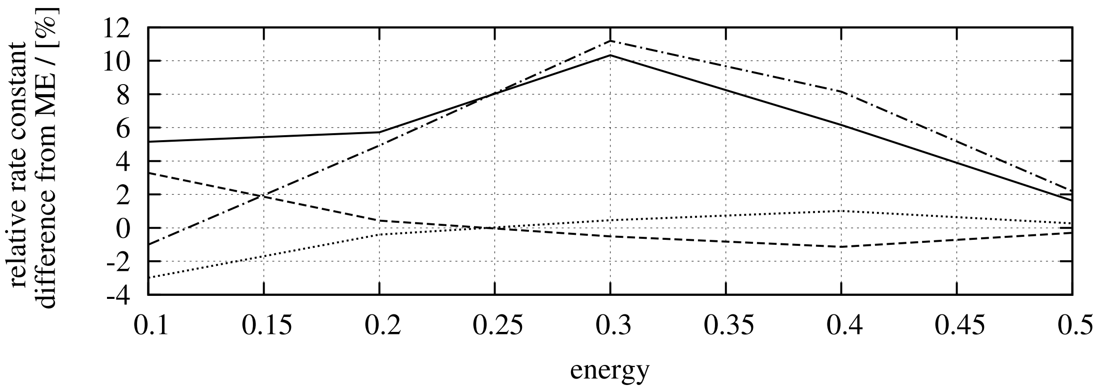

Figure 3.4: Comparison of rate constants obtained from FPT-BXD in boxes with potential (3.8). The relative difference in per cent is given as a function of energy. The statistical error of all values in the graph is roughly ±3%. Dashed and dotted lines represent the difference between ME rate constants (reference values) and FPT-BXD rate constants corrected with penetration times $k_{\mathrm{A}\to\mathrm{B}}$ and $k_{\mathrm{B}\to\mathrm{A}}$ using equation (2.14). The full and the dashed-dotted lines represent the difference between the ME rate constants and the exit FPT-BXD rate constants $k_{\mathrm{A}\to\partial\mathrm{AB}}$ and $k_{\mathrm{B}\to\partial\mathrm{AB}}$, respectively. The corrected FPT-BXD rate constants are in very good agreement with the reference ME values throughout the whole energy range.

artificially increases the population near box boundaries. Randomisation of the velocity direction after hitting the boundary, requiring the x component of velocity to point into the box, reproduces the correct distribution.

3.1.4 Higher Dimensions

To demonstrate that the above derived results are transferable to PES's more relevant for molecular systems, a many-dimensional potential of the form

$$\mathscr{V}(\mathbf{p}) = \cos(2\pi p_1) + \frac{1}{n}\sum_{i=1}^{n} a_n \sin(6\pi \mathbf{p}\cdot\mathbf{c}_i) \tag{3.9}$$

was studied, where the parameters a_i and $\mathbf{c}_i$ were selected as random numbers (uniform distribution) between 0 and 1. In this work, an $n = 7$-dimensional system was studied. The box was defined as $p_i \in [0, 1]$ for each i. The potential has on average two minima in each dimension per box, so the total number of minima can be estimated as 2^7. Simulations show that such a potential does not need randomisation lines and the population evolution is roughly exponential.

Boxes defined by Voronoi tesselation for complex molecular systems can sometimes lack the funnel-like structure modelled in potential (3.9) with the first term. Therefore, relaxation to equilibrium for the potential without the cosine term was also studied. The evolution (figure 3.5) can be decomposed into two exponentials.

A similar decomposition of species into more types was considered by Bunker and Hase[47;99;100] for different potentials. The decomposition might be used to describe systems divided into boxes with non-exponential kinetics more accurately in terms of the master equation.

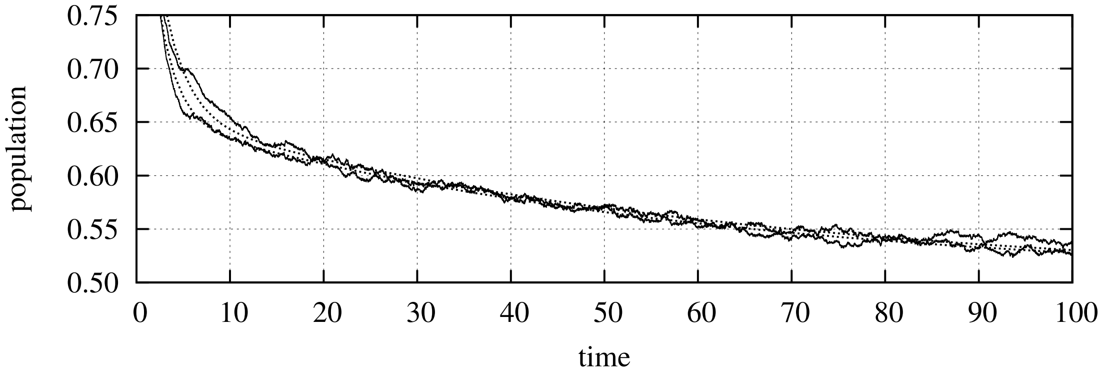

Figure 3.5: Decomposition of FPT distributions of passing into two different neighbouring boxes. Potential (3.9) was used. For energy = 0.2, the weight of the slowly decreasing exponential is 0.32 for neighbour 1 and 0.31 for neighbour 2. The similarity of the weights suggests that the box can be subdivided into two phase space boxes.

3.2 Simulation of a Cluster of Lennard-Jones Disks

3.2.1 Characterisation of the System

A cluster of seven Lennard-Jones (LJ) atoms in a two-dimensional plane ($\mathrm{LJ}_7^{\mathrm{2D}}$) provides a simple model suitable for testing new methods. The system consists of seven atoms of equal mass interacting with each other through the LJ pairwise potential,[101] having 11 internal degrees of freedom. The phase state of the system can be described by 14 coordinates: $\mathbf{q} = (x_1, y_1, ..., x_7, y_7)$, and 14 momenta $\mathbf{p} = (p_{x1}, p_{y1}, ..., p_{x7}, p_{y7})$. The Hamiltonian of the system is

$$\mathscr{H}(\mathbf{q},\mathbf{p}) = \frac{\mathbf{p}\cdot\mathbf{p}}{2m} + \sum_{i=1}^{6}\sum_{j=i+1}^{7} 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{3.10}$$

where $r_{ij} = \left((x_i - x_j)^2 + (y_i - y_j)^2\right)^{1/2}$ is the distance between particle i and j and m is the mass of each particle. σ (collision radius) and ϵ (well depth) are the parameters of the LJ potential defining the reduced units, such as the reduced time $\tau_0 = (m\sigma^2/\epsilon)^{1/2}$. The disconnectivity graph[4;102] of $\mathrm{LJ}_7^{\mathrm{2D}}$ is shown in figure 3.6.

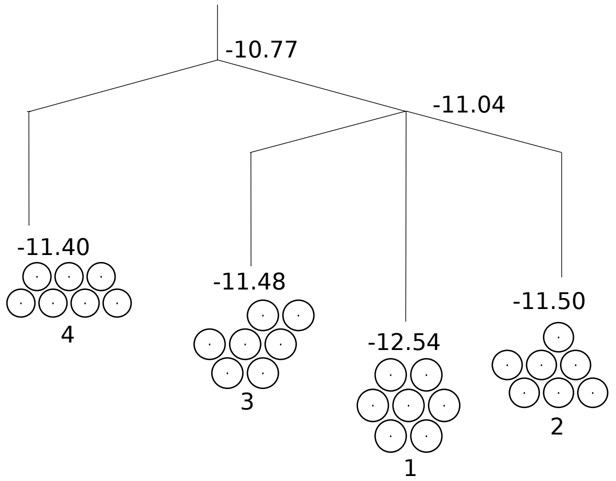

Figure 3.6: Minima and transition states of the LJ_7^{2D} system.[4;81] Energies are in reduced units (ϵ). The index of each minimum in order of increasing energy from the global minimum is given below each structure.

3.2.2 Implementation

Calculations using boxes defined by minima only gives a good value for the rate constant $k_{3\to2}$ (transition from minimum 3 to minimum 2, see figure 3.6) but the other rate constants, such as $k_{1\to2}$, are so low that the simulation would be too computationally expensive. Therefore, the box centres were instead defined as points on pathways in configuration space between the minima. The pathways were obtained using the doubly-nudged elastic band (DNEB) method[103] in the OPTIM[104] program. A modified L-BFGS minimiser[105] was used for optimisation of the positions of the images along the path. The points from each path were selected so that the energy difference between each two following points was always less than 0.6. Redundant points shared by more pathways or very close in energy and RMSD were removed. In the end, three different configuration space partitionings were used using 4, 13 and 22 boxes. The structures of the box centres for the partitioning using 22 boxes with their energies is shown in figure 3.7.

BXD simulations with boxes defined by Voronoi construction are highly dependent on a fast and reliable calculation of the “distance”. Out of the possible distance measures, RMSD is the most widely used. Finding the optimum structural alignment for RMSD calculation with respect to rotational and permutation isomers is a difficult computational problem. There are efficient algorithms for finding the optimum mutual orientation[106] for a given permutation and for finding the optimum permutation[107;108] for a given mutual rotation. However, the only known deterministic algorithm that ensures convergence to the global minimum of the RMSD with

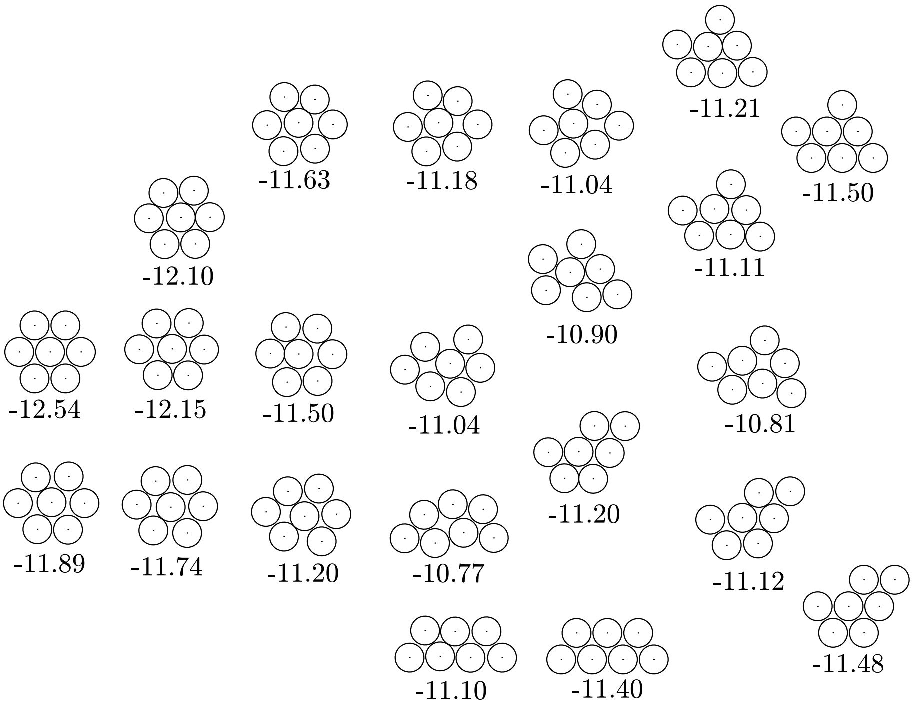

Figure 3.7: Structures representing generating points for boxes in the Voronoi construction. Energies are given below each structure in ϵ. Some structures are very close to the neighbouring box centres (RMSD$<$ 0.1) due to the steep potential and the requirement that energy differences of consecutive generating structures on any path cannot be higher than 0.5 ϵ.

respect to both types of alignment is search over all permutations. GMIN[109] calculates the RMSD's quickly by applying a modified Hungarian algorithm[108] first, followed by optimisation of the mutual rotation.

An alternative alignment method was used in this work. The centres of mass of two planar structures A and B are first positioned to the same point. Structure B is then rotated by an angle θ and the optimum permutation is found using the modified Hungarian algorithm.[108] For the resulting optimum permutation, the angle is optimised with the orthogonal transformation using quaternions.[106] The application of the Hungarian algorithm and the angle optimisation is performed for n uniformly distributed angles θ (from 0 to 2π). The method neither ensures convergence to the global minimum nor is computationally cheap. However, the premise that a systematic search through angles will find the global minimum is supported by a known solution to a similar problem.[110] The method was tested on a set of 489 structures representing paths between minima calculated using DNEB. The results for $n = 20$ were identical to the results obtained by the search over all permutations.

The BXD was implemented within the GMIN program[109] as a new procedure BXD2D. Nose-Hoover[111;112] and Berendsen[113] thermostats were used. The velocity inversion after hitting the wall was performed as in Glowacki *et al.*[5] The rate constants for the transition between particular boxes were calculated using the NGT procedure[87] implemented in PELE.[114]

3.2.3 Preliminary Results

The simulation protocol is currently not optimised to yield results comparable to TPS calculations.[92] Classical simulations in the canonical ensemble of $\mathrm{LJ}_7^{2\mathrm{D}}$ system were performed at $T = 0.05$ for evolution times of 10^4 τ_0. As mentioned above, the results for four boxes give values comparable to TPS for $k_{3\to 2}$, which is large enough to be sampled. Therefore, the system must be subdivided into more boxes. The rate constants calculated from simulations with 13 or 22 boxes are significantly higher than the TPS rate constants.

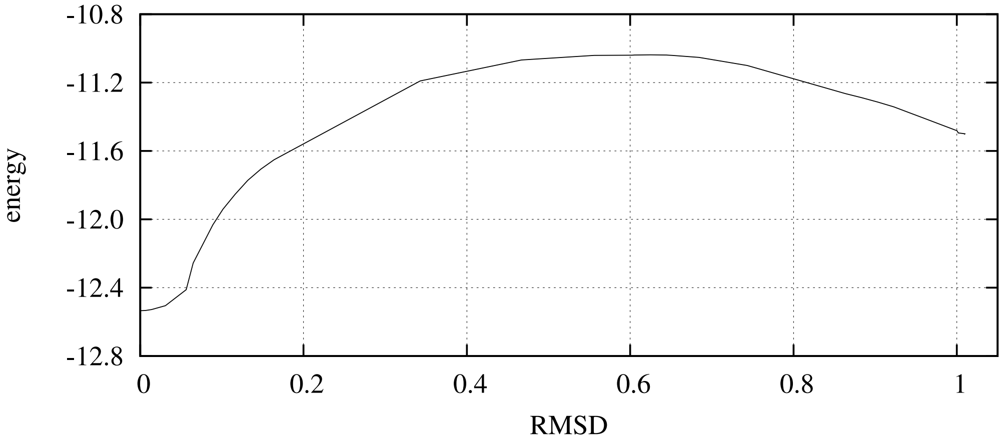

Figure 3.8: The energy profile of the transition from minimum 1 (the global minimum) to minimum 2. The RMSD between each point of the path and minimum 1 is calculated using the new method described in section 3.2.2.

The check for Markovianity proposed in section 2.6 showed that the input and output trajectories are strongly correlated in case of the partitioning into 22 boxes. First, the system is too simple to sufficiently decorrelate the input and output trajectories. It is obvious from figure 3.7 that the RMSD's between some box centres are very small. The energy increases very quickly with RMSD (5 ϵ/σ, see figure 3.8) along the transition path from minimum 1 to any other minimum. Second, as mentioned in section 2.6, the inversion procedure by Glowacki *et al.*[5] does not preserve the equilibrium distribution and significantly increases the number of the inversion events. Randomisation of velocities is more difficult to implement for this system. These problems will hopefully to be resolved in the near future.

Chapter 4

Conclusions and Future Work

In this work, a new formula for calculating rate constants by boxed molecular dynamics simulations is presented. The formula is based on the concept of first passage times defined in the phase space and makes use of the fact that each trajectory of non-zero time length contains an infinite number of phase points for averaging. A theoretical formalism is developed and it is shown that the present approach does not require the assumption of strictly exponential relaxation. Toy models were used to demonstrate that constraining does not disturb the equilibrium distribution within the box and that FPT-BXD provides good rate constants for modelling in terms of the master equation. The exact TST rate constants were shown to agree with the exact classical rate constants for high energy barriers. A method for estimating non-Markovian behaviour has been proposed. Other approaches for rate constant calculation might also benefit from the proposed sampling formula. For example, in forward flux sampling,[115;116] the MFPT can be used instead of the equilibrium (reactive flux) formulation for the calculation of the escape rate from the box representing the reactants.

The previous BXD method[5] can be used to accurately reproduce the thermodynamics of rare events if a proper velocity inversion procedure is employed, but the FPT-BXD must be used to simulate dynamics of the system. The space can be split into many boxes and the system can be simulated in terms of the master equation, or the rate constants of transitions between selected boxes can be calculated by a graph transformation method.[87] Apart from robustness with respect to dividing surface roughness, and properly including the effect of internal barriers, the newly developed method can be used for boxes with dividing surfaces that do not correspond to high energy barriers, which is often the case for physical systems of interest. The method also uses deterministic MD, so it does not depend on a phenomenological friction constant.

Preliminary results for FPT-BXD simulations of an $\mathrm{LJ}_7^{\mathrm{2D}}$ cluster with boxes defined by Voronoi construction show that small boxes can result in high correlation between the input and output trajectories. Steep energy barriers can make partitioning of the configuration space into boxes to provide efficient sampling difficult. In further studies, we plan to further benchmark and optimise the simulation protocol:

- The most important short-term aim is to develop an inversion procedure that correctly reproduces the equilibrium flux and MFPT's. The inversion procedure could be then also used to simulate small neighbourhoods of stationary points efficiently in order to correct the TST rate constants including anharmonicity. An entirely different initialisation procedure, such as using points from the simulation not located at the boundaries, can be used as a reference.

- More complicated toy models, such as the PES of a collinear atom transfer reaction,[117] can be used to gradually proceed from one particle in two dimensions to more realistic systems. Models with steep barriers and increasingly complex landscapes can be used to study the efficiency limits of the method.

- There is much scope for further theoretical development. Understanding how to divide the configuration space boxes into a larger number of phase space boxes with efficient determination of the rate constants using FPT-BXD would also broaden the applicability of the method.

- FPT-BXD can be coupled with hyperdynamics[74] in order to sample deep minima without partitioning the configuration space of system into more boxes along transition paths.

- Simulations of $\mathrm{LJ}_7^{\mathrm{2D}}$ have shown that the selection of the boxes affects the efficiency of the rate constant calculation. Boxes were defined by box centres used in Voronoi tesselation. In weighted Voronoi tesselation, the partitioning depends also on *ad hoc* parameter $w(i, j)$ [see equation (1.24)] defined for each pair of boxes. We currently use $w(i, j) \equiv 1$ for every pair. Automatic on-the-fly modification of this parameter in order to achieve comparable MFPT's in neighbouring boxes can increase efficiency of the algorithm. Box centres could be defined by a method similar to the one used by Chodera *et al.*[13] for large systems.

In future, we plan to apply the method to interesting systems:

- Application to three-dimensional clusters requires a reliable and fast structural alignment procedure. The currently implemented algorithm for 2D alignment can be generalised for finding the optimum permutation-rotations isomers in 3D. The golden section spiral algorithm can be used for finding evenly distributed points on a sphere.[118]

- The alanine dimer could serve as a good system for benchmarking the method. Trp-cage and other similar small peptides can be used as test systems and then larger proteins can be studied.

- The dynamics of large proteins can be simulated with FPT-BXD, perhaps in combination with some recently developed rigidification algorithms.[119;120] The dynamics of protein conformational transitions and protein folding are being studied extensively[121;122] and FPT-BXD has the potential to contribute.

References

[1] D. J. Wales and H. A. Scheraga, *Science*, 285; 1368–1372, **1999**.

[2] M. Paterson and T. Przytycka, *Discret Appl. Math.*, 71; 217–230, **1996**.

[3] P. Crescenzi, D. Goldman, C. Papadimitriou, A. Piccolboni and M. Yannakakis, *J. Comput. Biol.*, 5; 423–465, **1998**.

[4] D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge, **2003**.

[5] D. R. Glowacki, E. Paci and D. V. Shalashilin, *J. Phys. Chem. B*, 113; 16603–16611, **2009**.

[6] M. Holodniok, A. Klíč, M. Kubíček and M. Marek, *Metody analýzy nelineárních dynamických modelů*, Academia, Praha, **1986**.

[7] D. T. Gillespie, *Physica A*, 188; 404–425, **1992**.

[8] V. S. Pande, K. Beauchamp and G. R. Bowman, *Methods*, 52; 99–105, **2010**.

[9] P. M. Kasson, N. W. Kelley, N. Singhal, M. Vrljic, A. T. Brunger and V. S. Pande, *Proc. Natl. Acad. Sci. USA*, 103; 11916–11921, **2006**.

[10] N. W. Kelley, V. Vishal, G. A. Krafft and V. S. Pande, *J. Chem. Phys.*, 129, **2008**.

[11] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, **1992**.

[12] J. D. Chodera, "Master equation models of macromolecular dynamics from atomistic simulation", Ph.D. thesis, UCSF, **2006**.

[13] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill and W. C. Swope, *J. Chem. Phys.*, 126, **2007**.

[14] J. D. Chodera, W. C. Swope, J. W. Pitera and K. A. Dill, *Multiscale Model. Simul.*, 5; 1214–1226, **2006**.

[15] F. Noe and S. Fischer, *Curr. Opin. Struct. Biol.*, 18; 154–162, **2008**.

[16] G. R. Bowman, X. Huang and V. S. Pande, *Methods*, 49; 197–201, **2009**.

[17] G. R. Bowman, K. A. Beauchamp, G. Boxer and V. S. Pande, *J. Chem. Phys.*, 131; 124101, **2009**.

[18] J. H. van't Hoff, *Etudes de dynamique chimique*, Amsterdam, **1884**.

[19] S. Arrhenius, *Z. Physik. Chem.*, 4; 226, **1889**.

[20] L. Farkas, *Z. Physik. Chem.*, 125; 236–242, **1927**.

[21] H. Eyring, *J. Chem. Phys.*, 3; 107–115, **1935**.

[22] M. Polanyi, *Z. Phys.*, 2; 90–110, **1920**.

[23] E. Wigner, *J. Chem. Phys.*, 5; 720–723, **1937**.

[24] E. Wigner, *J. Chem. Phys.*, 7; 646–650, **1939**.

[25] E. Wigner, *Trans. Faraday Soc.*, 34; 0029–0040, **1938**.

[26] M. Born and R. Oppenheimer, *Ann. Phys.-Berlin*, 84; 0457–0484, **1927**.

[27] O. K. Rice and H. C. Ramsperger, *J. Am. Chem. Soc.*, 49; 1617–1629, **1927**.

[28] O. K. Rice and H. C. Ramsperger, *J. Am. Chem. Soc.*, 50; 617–620, **1928**.

[29] L. S. Kassel, *J. Phys. Chem.*, 32; 225–242, **1928**.

[30] R. A. Marcus, *J. Chem. Phys.*, 24; 966–978, **1956**.

[31] D. Chandler, *J. Chem. Phys.*, 68; 2959, **1978**.

[32] R. Kubo, *J. Phys. Soc. Jpn.*, 12; 570–586, **1957**.

[33] L. Onsager, *Phys. Rev.*, 37; 405–426, **1931**.

[34] L. Onsager, *Phys. Rev.*, 38; 2265–2279, **1931**.

[35] D. Frenkel and B. Smit, *Understanding Molecular Simulations: from Algorithms to Applications*, Academic Press, San Diego, **2002**.

[36] D. G. Truhlar and B. C. Garrett, *Accounts Chem. Res.*, 13; 440–448, **1980**.

[37] P. L. Fast and D. G. Truhlar, *J. Chem. Phys.*, 109; 3721–3729, **1998**.

[38] L. R. Pratt, *J. Chem. Phys.*, 85; 5045–5048, **1986**.

[39] T. Uzer, C. Jaffe, J. Palacian, P. Yanguas and S. Wiggins, *Nonlinearity*, 15; 957–992, **2002**.

[40] G. S. Ezra, H. Waalkens and S. Wiggins, *J. Chem. Phys.*, 130, **2009**.

[41] H. A. Kramers, *Physica*, 7; 284–304, **1940**.

[42] R. F. Grote and J. T. Hynes, *J. Chem. Phys.*, 73; 2715–2732, **1980**.

[43] E. Pollak, *J. Chem. Phys.*, 85; 865–867, **1986**.

[44] D. G. Truhlar, B. C. Garrett and S. J. Klippenstein, *J. Phys. Chem.*, 100; 12771–12800, **1996**.

[45] E. Pollak and P. Talkner, *Chaos*, 15; 47, **2005**.

[46] E. Vanden-Eijnden and F. A. Tal, *J. Chem. Phys.*, 123, **2005**.

[47] D. L. Bunker and W. L. Hase, *J. Chem. Phys.*, 59; 4621–4632, **1973**.

[48] U. Lourderaj and W. L. Hase, *J. Phys. Chem. A*, 113; 2236–2253, **2009**.

[49] P. Hanggi, P. Talkner and M. Borkovec, *Rev. Mod. Phys.*, 62; 251–341, **1990**.

[50] R. Muller, P. Talkner and P. Reimann, *Physica A*, 247; 338–356, **1997**.

[51] J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 23; 327–341, **1977**.

[52] M. Tuckerman, B. J. Berne and G. J. Martyna, *J. Chem. Phys.*, 97; 1990–2001, **1992**.

[53] A. F. Voter, *Phys. Rev. B*, 57; 13985–13988, **1998**.

[54] R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.*, 57; 2607–2609, **1986**.

[55] A. Laio and M. Parrinello, *Proc. Nat. Acad. Sci. USA*, 99; 12562–12576, **2002**.

[56] W. E, W. Ren and E. Vanden-Eijnden, page arXiv:math/0212415, **2002**.

[57] C. Chipot and A. Pohorille (Editors), *Free Energy Calculations*, Springer, **2007**.

[58] C. Dellago and P. G. Bolhuis, *Adv. Polym. Sci.*, 221; 167–233, **2009**.

[59] M. A. Rohrdanz, W. Zheng and C. Clementi, *J. Chem. Phys.*, 138; 164112, **2013**.

[60] D. Moroni, "Efficient sampling of rare event pathways", Ph.D. thesis, Universiteit van Amsterdam, **2005**.

[61] C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.*, 108; 1964–1977, **1998**.

[62] C. Dellago, P. G. Bolhuis and D. Chandler, *J. Chem. Phys.*, 110; 6617–6625, **1999**.

[63] P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 53; 291–318, **2002**.

[64] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 21; 1087–1092, **1953**.

[65] P. G. Bolhuis, C. Dellago and D. Chandler, *Farad. Discuss.*, 110; 421–436, **1998**.

[66] D. Moroni, T. S. van Erp and P. G. Bolhuis, *Phys. Rev. E*, 71; 056709, **2005**.

[67] E. E. Borrero and C. Dellago, *J. Chem. Phys.*, 133, **2010**.

[68] E. A. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, *Chem. Phys. Lett.*, 156; 472–477, **1989**.

[69] E. Paci, G. Ciccotti, M. Ferrario and R. Kapral, *Chem. Phys. Lett.*, 176; 581–587, **1991**.

[70] E. Paci and G. Ciccotti, *J. Phys. Condens. Matter*, 4; 2173–2184, **1992**.

[71] M. Sprik and G. Ciccotti, *J. Chem. Phys.*, 109; 7737–7744, **1998**.

[72] G. Ciccotti, R. Kapral and E. Vanden-Eijnden, *ChemPhysChem*, 6; 1809–1814, **2005**.

[73] E. Martinez-Nunez and D. V. Shalashilin, *J. Chem. Theory Comput.*, 2; 912–919, **2006**.

[74] A. F. Voter, *J. Chem. Phys.*, 106; 4665, **1997**.

[75] D. Hamelberg, J. Mongan and J. McCammon, *J. Chem. Phys.*, 120; 11919–11929, **2004**.

[76] U. Doshi and D. Hamelberg, *J. Chem. Theory Comput.*, 8; 4004–4012, **2012**.

[77] M. R. Sorensen and A. F. Voter, *J. Chem. Phys.*, 112; 9599–9606, **2000**.

[78] A. K. Faradjian and R. Elber, *J. Chem. Phys.*, 120; 10880–10889, **2004**.

[79] E. Vanden-Eijnden, M. Venturoli, G. Ciccotti and R. Elber, *J. Chem. Phys.*, 129; 174102, **2008**.

[80] E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.*, 130; 194101, **2009**.

[81] D. Wales, *Mol. Phys.*, 100; 3285–3305, **2002**.

[82] D. J. Wales, *Int. Rev. Phys. Chem.*, 25; 237–282, **2006**.

[83] J. M. Carr and D. J. Wales, *Phys. Chem. Chem. Phys.*, 11; 3341–3354, **2009**.

[84] J. Pancíř, *Collect. Czech. Chem. Commun.*, 40; 1112–1118, **1975**.

[85] C. J. Cerjan and W. H. Miller, *J. Chem. Phys.*, 75; 2800–2806, **1981**.

[86] G. Henkelman, B. P. Uberuaga and H. Jonsson, *J. Chem. Phys.*, 113; 9901–9904, **2000**.

[87] D. J. Wales, *J. Chem. Phys.*, 130, **2009**.

[88] Y. Guo, D. V. Shalashilin, J. A. Krouse and D. L. Thompson, *J. Chem. Phys.*, 110; 5514–5520, **1999**.

[89] D. V. Shalashilin and D. L. Thompson, *J. Chem. Phys.*, 107; 6204–6212, **1997**.

[90] D. V. Shalashilin, G. S. Beddard, E. Paci and D. R. Glowacki, *J. Chem. Phys.*, 137, **2012**.

[91] D. R. Glowacki, E. Paci and D. V. Shalashilin, *J. Chem. Theory Comput.*, 7; 1244–1252, **2011**.

[92] C. Dellago, P. Bolhuis and D. Chandler, *J. Chem. Phys.*, 108; 9236–9245, **1998**.

[93] W. H. Miller, *Accounts Chem. Res.*, 9; 306–312, **1976**.

[94] A. B. Bortz, M. H. Kalos and J. L. Lebowitz, *J. Comput. Phys.*, 17; 10–18, **1975**.

[95] D. T. Gillespie, *J. Comput. Phys.*, 22; 403–434, **1976**.

[96] K. A. Fichthorn and W. H. Weinberg, *J. Chem. Phys.*, 95; 1090–1096, **1991**.

[97] L. Verlet, *Phys. Rev.*, 159; 98, **1967**.

[98] Octave community, “GNU/Octave”, **2012**. URL `www.gnu.org/software/octave`

[99] D. L. Bunker, *J. Chem. Phys.*, 40; 1946, **1964**.

[100] W. L. Hase, D. G. Buckowski and K. N. Swamy, *J. Phys. Chem.*, 87; 2754–2763, **1983**.

[101] J. E. Lennard-Jones and A. E. Ingham, *Proc. R. soc. Lond. Ser. A*, 107; 636–653, **1925**, -Contain. Pap. Math. Phys. Character.

[102] O. M. Becker and M. Karplus, *J. Chem. Phys.*, 106; 1495–1517, **1997**.

[103] S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.*, 120; 2082, **2004**.

[104] Wales, D. J. , “OPTIM: a program for optimizing geometries and calculating reaction pathways”, **2013**. URL `www-wales.ch.cam.ac.uk/OPTIM`

[105] J. Nocedal, *Math. Comput.*, 35; 773–782, **1980**.

[106] S. K. Kearsley, *Acta Crystallogr. Sect. A*, 45; 208–210, **1989**.

[107] H. W. Kuhn, *Nav. Res. Logis. Quart.*, 2; 8397, **1955**.

[108] R. Jonker and A. Volgenant, *Computing*, 38; 325–340, **1987**.

[109] Wales, D. J. , “GMIN: A program for basin-hopping global optimisation”, **2013**. URL `www-wales.ch.cam.ac.uk/GMIN`

[110] B. Helmich and M. Sierka, *J. Comput. Chem.*, 33; 134–140, **2012**.

[111] S. Nosé, *J. Chem. Phys.*, 81; 511–519, **1984**.

[112] D. S. Kleinerman, C. Czaplewski, A. Liwo and H. A. Scheraga, *J. Chem. Phys.*, 128, **2008**.

[113] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola and J. Haak, *J. Chem. Phys.*, 81; 3684–3690, **1984**.

[114] Stevenson, Jacob and Ruehle, Victor and Wales, D. J. and pele community, “PELE : Python Energy Landscape Explorer”, **2013**. URL `www.github.com/pele-python/pele`

[115] R. J. Allen, C. Valeriani and P. R. ten Wolde, *J. Phys. Condens. Matter*, 21, **2009**.

[116] K. Kratzer, A. Arnold and R. J. Allen, *J. Chem. Phys.*, 138, **2013**.

[117] D. Secrest and R. Johnson, *J. Chem. Phys.*, 45; 4556, **1966**.

[118] E. B. Saff and A. B. J. Kuijlaars, *Math. Intell.*, 19; 5–11, **1997**.

[119] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das and H. C. Andersen, *J. Chem. Phys.*, 128, **2008**.

[120] H. Kusumaatmaja, C. S. Whittleston and D. J. Wales, *J. Chem. Theory Comput.*, 8; 5159–5165, **2012**.

[121] R. B. Best, *Curr. Opin. Struct. Biol.*, 22; 52–61, **2012**.

[122] K. A. Dill and J. L. MacCallum, *Science*, 338; 1042–1046, **2012**.