MOLECULAR DYNAMICS AND STOCHASTIC SIMULATIONS OF SURFACE DIFFUSION

A Thesis
Presented to
The Academic Faculty

by

Jeremy Moix

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
May 2007

MOLECULAR DYNAMICS AND STOCHASTIC SIMULATIONS OF SURFACE DIFFUSION

Approved by:

Dr. Rigoberto Hernandez, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Thomas Orlando
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Peter J. Ludovice
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Jean-Luc Brèdas
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. C. David Sherrill
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: 21 March 2007

ACKNOWLEDGEMENTS

This work would not have been completed had it not been for several figures who served as a constant source of support and encouragement. First and foremost, I would like to thank my advisor, Professor Rig Hernandez, who has been my most steadfast supporter and has exceeded all of my expectations. His guidance has been instrumental to my success and I will be forever grateful. Additionally, I would like to thank my coworkers and officemates in the Sherrill and Hernandez groups with whom I have enjoyed many discussions throughout the years over beers, coffee and Little Debbies®: Shi Zhong, Berhane Temelso, John Sears, Yanping Qin, Ashley Ringer, Ashley Tucker, Steve Arnstein, Gungor Ozer, Jay Foley. I would also like to thank Dr. Alex Popov who has of late provided much support and instruction with little real benefit to himself. Furthermore, I would like to acknowledge all of my collaborators particularly, Professor Turgay Uzer and Dr. Thomas Bartsch who have generously and patiently shared their pearls of wisdom even though they rarely received any in return. I would also like to thank Professors Ed Valeev and Micah Abrams who took me under their wing during my first years as a graduate student and provided of an enormous amount of guidance and inspiration. And finally none of this would have been possible without the constant support of my family and friends. Their encouragement over the years has been invaluable. Thank you all.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES vii

LIST OF FIGURES viii

SUMMARY x

I INTRODUCTION 1

- 1.1 Classical Surface Diffusion 4
- 1.2 Molecular Dynamics Simulations and Projective Models 8
  - 1.2.1 Atomistic Models 8
  - 1.2.2 Projective Models 10
- 1.3 Overview 13

II REDUCED DIMENSIONAL MODELS DERIVED FROM MOLECULAR DYNAMICS SIMULATIONS OF TRANSPORT ON METAL SURFACES 16

- 2.1 Introduction 16
- 2.2 Molecular Dynamics Simulations 18
  - 2.2.1 Diffusion Rates 19
- 2.3 Reduced-Dimensional Systems 20
- 2.4 Conclusions 23

III A PHENOMENOLOGICAL MODEL FOR SURFACE DIFFUSION: DIFFUSIVE DYNAMICS ACROSS STOCHASTIC POTENTIALS 25

- 3.1 Introduction 25
- 3.2 Model and Methods 27
- 3.3 Results and Discussion 31
- 3.4 Concluding Remarks 38

IV DISSIPATING THE LANGEVIN EQUATION IN THE PRESENCE OF AN EXTERNAL STOCHASTIC POTENTIAL 40

- 4.1 Introduction 40
- 4.2 Langevin Model with Stochastic Potentials 42
  - 4.2.1 Stochastic Potential Representation 42
  - 4.2.2 Uniform Dissipation 43

4.2.3 Space-Dependent Dissipation 44
4.3 Results and Discussion 46
4.4 Conclusions 53

V IDENTIFYING REACTIVE TRAJECTORIES USING A MOVING TRANSITION STATE 55
5.1 Introduction 55
5.2 Preliminaries 57
5.2.1 The Two-Dimensional Dissipative Model 58
5.2.2 The Transition State Trajectory 60
5.2.3 The Relative Dynamics 61
5.3 The Barrier Ensemble 62
5.4 Analytic Results 64
5.4.1 Reaction Probabilities 65
5.4.2 Reaction Times 66
5.5 Numerical Results 69
5.5.1 Harmonic Systems 70
5.5.2 Nonlinear Systems 76
5.6 Concluding Remarks 81

VI DIHEDRAL-ANGLE INFORMATION ENTROPY AS A GAUGE OF SECONDARY STRUCTURE PROPENSITY 84
6.1 Introduction 84
6.2 Methods 86
6.2.1 ϕ_i-ψ_i and ψ_i-ϕ_{i+1} Distributions 86
6.2.2 Data-mining the ψ_i-ϕ_{i+1} Distributions 87
6.2.3 The Dihedral-angle Information Entropy 89
6.2.4 A Checking Function for Secondary Structure Propensity 91
6.3 Results and Discussion 93
6.3.1 Dihedral Angle Distributions 93
6.3.2 On the Choice of the Sequence Database 93
6.3.3 D_1 and D_2 Checks 96
6.3.4 The Role of D_2 in Checking Theoretical Structures 99

6.3.5 D_2 and Other Checking Functions 102
6.4 D_2Check Web Server 103
6.5 Conclusion 104
VII CONCLUSIONS 107
APPENDIX A PROPERTIES OF THE STOCHASTIC BARRIERS 112
APPENDIX B SPACE-DEPENDENT FRICTION FOR MERGED HARMONIC OSCILLATOR POTENTIALS 114
APPENDIX C BACKWARD TIME STOCHASTIC INTEGRATOR 117
APPENDIX D CONSTRUCTION OF THE DIHEDRAL-ANGLE DISTRIBUTIONS 119
REFERENCES 126

LIST OF TABLES

1 Corrections to the thermal friction constant 48
2 Friction corrections for sinusoidal potentials 49
3 Protein structures contained in each library 88
4 Atypical theoretical and experimental structures in PDB 100
5 D_2 values for several template and model protein structures 101
6 D_2 and other checking function scores for several proteins 102
7 Relative abundance of all possible residue pairs 123

LIST OF FIGURES

1 Prototypical adsorbed molecular systems on metal surfaces 2

2 Temperature dependence of diffusion coefficients 20

3 Dimensionless diffusion profiles as a function of temperature 21

4 Potential of mean force for the Ag (1 0 0) unit cell 22

5 Potential of mean force along the diffusion coordinate 22

6 Force autocorrelations evaluated at the well and barrier 23

7 Mean first passage times at high friction 31

8 Mean first passage times at low friction 33

9 Mean first passage times at large variance 34

10 Normalized mean first passage times 35

11 Rate enhancements for one-dimensional stochastic potentials 37

12 Rate enhancements for coherent two-dimensional stochastic potentials . . . 37

13 Rate enhancements for incoherent two-dimensional stochastic potentials . . 38

14 Space-dependent friction for merged harmonic oscillator potentials 46

15 Space-dependent friction for sinusoidal potentials 47

16 Mean first passage times with uniform and space-dependent dissipation on incoherent potentials at small variance 51

17 Mean first passage times with uniform and space-dependent dissipation on incoherent potentials at large variance 51

18 Mean first passage times with uniform and space-dependent dissipation on coherent potentials 52

19 Qualitative dynamics of the barrier ensemble in relative coordinates 63

20 Reaction time distributions 68

21 Illustrative realization the transition state trajectory and barrier ensemble . 71

22 Reaction probabilities with various dividing surfaces 71

23 Average recrossings of the transition state in nonlinear systems 72

24 Numerical reaction probabilities for a given instance of the noise 73

25 Average reaction probabilities 75

26 Reaction probability distributions 75

27 Reaction probabilities in nonlinear systems 77

28 Average recrossings of the dividing surfaces 78
29 Fraction of correctly identified trajectories 80
30 Average reaction probabilities in nonlinear systems 81
31 Illustrative tripeptide dihedral angles 87
32 Distributions of the ψ and ϕ dihedral angles 89
33 Structure entropy distributions 94
34 Distributions of ΔS evaluated for various libraries 95
35 D_1 distribution evaluated for various libraries 97
36 D_2 distribution evaluated for various libraries 97
37 D_2Check server output 104
38 ψ_i-ϕ_{i+1} probability distributions 122
39 Correlation between ψ_i and ϕ_{i+1} 124

SUMMARY

Despite numerous advances in experimental methodologies capable of addressing the various phenomenon occurring on metal surfaces, atomic scale resolution of the microscopic dynamics remains elusive for most systems. Computational models of the processes may serve as an alternative tool to fill this void. To this end, parallel molecular dynamics simulations of self-diffusion on metal surfaces have been developed and employed to address microscopic details of the system. However these simulations are not without their limitations and prove to be computationally impractical for a variety of chemically relevant systems, particularly for diffusive events occurring in the low temperature regime. To circumvent this difficulty, a corresponding coarse-grained representation of the surface is also developed resulting in a reduction of the required computational effort by several orders of magnitude, and this description becomes all the more advantageous with increasing system size and complexity. This representation provides a convenient framework to address fundamental aspects of diffusion in nonequilibrium environments and an interesting mechanism for directing diffusive motion along the surface is explored. In the ensuing discussion, additional topics including transition state theory in noisy systems and the construction of a checking function for protein structure validation are outlined. For decades the former has served as a cornerstone for estimates of chemical reaction rates. However, in complex environments transition state theory most always provides only an upper bound for the true rate. An alternative approach is described that may alleviate some of the difficulties associated with this problem. Finally, one of the grand challenges facing the computational sciences is to develop methods capable of reconstructing protein structure based solely on readily-available sequence information. Herein a checking function is developed that may prove useful for addressing whether a particular proposed structure is a viable possibility.

CHAPTER I

INTRODUCTION

The motion of adsorbates on metal surfaces remains an area of active research as a result of its technological implications for a diverse set of processes such a catalysis, epitaxial growth, and self-assembly, among many others. Aside from the relevance to practical applications, fundamental scientific interest remains strong because of the rich physics governing these processes and the challenges facing both the experimental and theoretical communities to develop accurate techniques that are capable of investigating and describing the various phenomenon. Nevertheless, these communities have made great advances over the past two decades. The curious reader is directed to several excellent reviews of recent contributions in experimental, theoretical and computational methodologies, and their interpretation from experimental [1, 2, 3, 4, 5, 6] or theoretical [7, 8, 9, 10, 11, 12] perspectives, although obviously neither can be completely devoid of the other. For a review discussing the challenges and recent successes in modeling surface dynamics across very large length and time scales while retaining atomic level details as needed, see reference [13]. Alternatively, algorithmic developments attempting to accelerate molecular dynamics simulations of surface diffusion are reviewed in reference [14].

One of the primary goals of the work described here is to demonstrate that classical simulations of adsorbate dynamics on surfaces can be described by reduced-dimensional models and that these models can provide insight into their qualitative —and sometimes quantitative— behavior. In principle, one would prefer not to resort to coarse-grained models in favor of the inclusion of all degrees of freedom explicitly. However, such models remain computationally infeasible for many interesting chemical systems given the current state of computing capabilities and will remain so for some time in the near future given the foreseeable performance increases. Aside from their affordability, reduced-dimensional

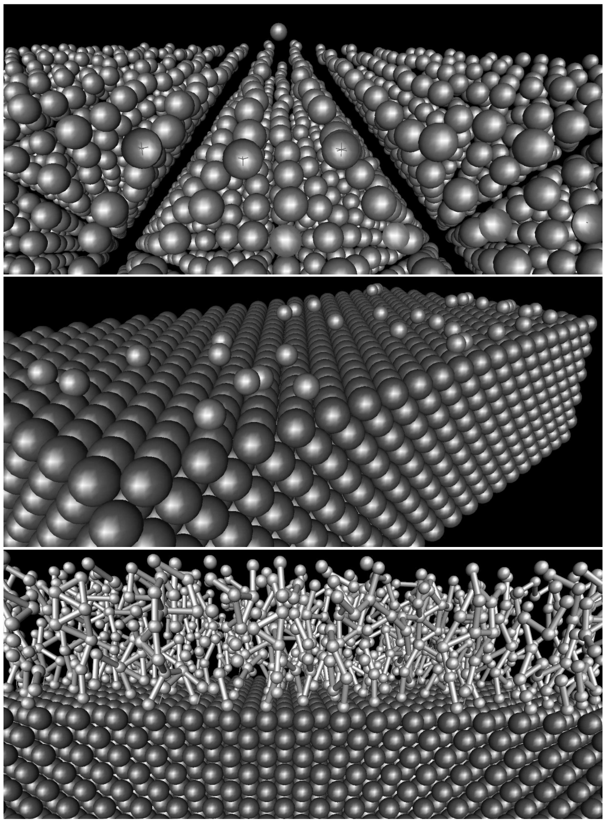

Figure 1: Much of the method development discussed here is focused on the equations of motion for a single particle on a surface as illustrated in the top graphic. However, the projective methods become all the more useful as one begins to consider many adsorbates (as in the middle graphic) or polymers (as in the bottom graphic). If the atomistic dynamics of the underlying surface can be subsumed through a single effective stochastic representation, then the dissipative MD can become tractable even for very large surface structures.

models also provide a lens on which modes of a system are important to the detailed dynamics as well as those that can be successfully ignored and are evidently not important. For example, in the cases illustrated in Figure 1, the all-atom simulations require the integration of the dynamics not just for the atoms, molecules, or polymers adsorbed on the surface, but of the much larger underlying surface. As the number and complexity of the adsorbates grows, it becomes increasingly useful to be able to ignore the detailed dynamics of the underlying surface, and this necessarily requires a projection. The latter is not a rigorous coarse-graining because the length scale of the remaining adsorbates has not been changed. However, these length scales may also need to be coarse-grained in order to assess macroscopic properties of the surface [13].

There is, of course, one very important additional level of complexity in these systems that arises when the length and energy scales are such that classical mechanics no longer offers the correct equations of motion. (The electronic dynamics are necessarily such a case.) Consequently, the correct inclusion of quantum mechanical degrees of freedom into the description of surface events and molecular simulations in general remains a topic at the forefront of current theoretical efforts. Unfortunately, such simulations are prohibitively expensive for processes occurring on all but the shortest of length and time scales. For reviews on the subject within the context of surface science one is referred to references [15, 16, 17, 10, 18]. While the methodology employed here will not focus on such approaches, it is perhaps worthwhile to point out that one potential advantage of the identification of reduced-dimensional models lies in the possibility that the subsystem can be quantized independently of the projected classical models [13, 19].

Presently simulations represent a compromise between the level of detail one would like to include and that which is capable of reasonably being included. This is particularly relevant when long-time correlations exist between the adsorbate and the surface. In this context, long-time scale implies long on the molecular scale but it should be noted this is usually very short on the experimental scale. The use of simulation in this instance has been instrumental in interpreting the complicated experimental results that typically represent an ensemble average over many adsorbates completing multiple hops between observations. One could certainly include ipso facto the long-time correlations through the use of an all-atom model that is sufficiently large as to include all of the requisite local modes, surface phonons, and bulk phonons while maintaining this classical-like prescription. However, the key aim here is the use of reduced-dimensional models that allow for the inclusion of the long-time correlations without recourse to such detailed models. Meanwhile such reduced-dimensional models, while projecting the bulk modes into dissipative terms can also project the electronic modes into the so-called electronic friction [20, 21, 10, 22]. Thus, in limits where the former has been seen to be effective, then the models described here with the appropriate friction terms can also account for quantum electronic effects. An essential requirement of any reduced-dimensional approach is that some transferability of the coarse-grained description will exist over the relevant parameter range of interest. Fortunately, this requirement is often satisfied. Obviously, one should not project out degrees of freedom of the system that will be important to the process of interest. However, once one has an acceptable model for surface diffusion including the appropriate long-time correlations then it can presumably be used to study various questions concerned with surface transport.

It is perhaps also helpful to briefly note a few experimental techniques in surface science and their relationship to the corresponding microscopic dynamics as elucidated through an appropriate reduced-dimensional or atomistic model. This duality can be exploited to assist and validate experimental or computational interpretations of complicated systems, and several reviews have discussed this topic in detail [1, 23, 3, 4, 5, 6, 11, 12]. For slow diffusion and low coverage, scanning tunneling microscopy (STM) and field ion microscopy (FIM) are able to directly follow diffusive events and thus provide a description of the microscopic time-dependent probability distribution [24, 25, 2, 26]. However, in more general cases, these methods lack the temporal resolution needed to observe faster hopping rates and it also becomes difficult to unambiguously identify the unique diffusion events as coverage increases. In such cases, quasi-elastic helium atom scattering (QHAS) becomes the preferred method although it provides only an indirect measure of adsorbate diffusion [23, 27, 28, 29]. The measured structure factor from these experiments is related to the microscopic hopping rates as determined for example, from a microscopic master equation.

1.1 Classical Surface Diffusion

Surface diffusion consists of a series of jumps between minima on the energy landscape. Each of these jumps is inherently a rare event (on the molecular scale) due to the large barriers impeding the hopping process which are typically on the order of a few tenths of an eV and generally an order of magnitude larger than the thermal energy [30, 31]. At relevant temperatures, such activated processes occur on time scales that lend themselves only to computationally-taxing classical molecular dynamics simulations or more feasible coarse-grained Langevin simulations. Taking into account that this is an activated process, the diffusion rate can usually be well represented by an Arrhenius-like form

$$D(T) = D_0(T) \exp\left(-\beta E_{\mathrm{A}}\right) \ . \tag{1}$$

where E_{A} is the classical activation energy, $\beta = 1/k_{\mathrm{B}}T$, and $D_0(T)$ is the pre-exponential factor whose temperature dependence is often weak [32]. This formalism provides one with the ability to extract the activation energy and prefactor from a series of simulations performed over different temperatures. However the simple interpretation is not quite adequate. At normal temperatures, an adsorbate would never have enough energy to surmount a barrier since typically $E_{\mathrm{A}} \gg k_{\mathrm{B}}T$. This elementary analysis makes it clear that the surface is a vital component of the diffusion process serving as an effective bath that is constantly exchanging energy with the adsorbates. The particle surmounts a barrier when it has accumulated a sufficient amount of energy from the surface and continues to diffuse across the barriers until the energy has been transferred back. This type of argument readily gives rise to a Langevin description of surface diffusion and has been invoked frequently in the past [33, 34, 35, 36, 37, 38]. The possibility of correlated hops to sites farther than the nearest wells during an excursion depends on the rate of energy relaxation. Because the probability of such events is usually low, correlated hops have often been ignored or included implicitly. However when the temperature is large or the adsorbate is weakly bound, then correlated hops have been seen to play an important role [37, 39, 40].

While the diffusion constant is dominated by the equilibrium barrier height E_{A}, all of the dynamical contributions are contained in the prefactor. Unfortunately, brute force atomistic simulations have only been computationally feasible for relatively high temperatures. (Of course, this constraint is constantly relaxing.) Molecular dynamics simulations are now possible for long runs of hundreds of nanoseconds or alternatively many short runs of several hundred picoseconds. However, in the case that the ratio of the barrier height to the thermal energy is large, acquiring sufficient hopping statistics through either approach becomes impractical at an exponential rate. As a result, one often takes advantage of the Arrhenius behavior and studies simulations at higher temperatures and then relies on an extrapolation to lower temperature regimes. In light of this difficulty alternative methods such as Monte

Carlo simulations are often preferred [8, 30, 41, 42, 43]. In this situation one relies on the assumption that hops are rare and uncorrelated events, and hence the entire trajectory can be taken as a series of independent (Markovian) processes. The diffusion coefficient may thus be cast as the product of a rate constant based on (canonical) transition state theory (TST) and the length of the hop (lattice constant), l. That is,

$$D = \frac{1}{2\alpha} k_{TST} l^2 , \tag{2}$$

provided all transition paths are identical where α represents the dimensionality of the system. Within this approximation, the elegant techniques developed from transition state theory may be employed thereby allowing one to extract very accurate rates with substantial computational savings [44, 45, 46, 47, 48]. However, the Markovian approximation of independent hops has recently been tested in the self-diffusion of Cu on Cu(1 1 1) and it was found that correlations exist between hops even at low temperatures when the assumption is expected to be valid [49]. This raises serious questions about blind applications of the Monte Carlo average of canonical TST rates, but clearly more work is needed given its success thus far.

Further caution is warranted given that not all of the elementary mechanisms comprising surface diffusion are fully understood. Traditional hops to nearest neighbors may be accompanied by many other processes including site exchange whereby the adsorbate and a surface atom exchange positions resulting in a net displacement, [50, 51] long jumps to sites further than nearest-neighbors, and possibly even sub-surface diffusion. However, these various complex mechanisms make reactive flux approaches difficult to employ when the transition path is not known or perhaps contains multiple geometrically disparate exit channels.

The situation is compounded by the fact that in order to achieve an accurate description of the bulk, many layers of surface atoms must be included in simulations [38]. In principle, the only way to know when one has included a large enough underlying surface is to continue to increase the number of layers until a clear convergence of the target observable has been reached. To circumvent this costly exercise, several investigations have allowed only those atoms within the vicinity of the adsorbate to move while others are pinned to their equilibrium positions [52, 53, 33, 34, 54]. While this results in a dramatic computational savings, it overly dampens the long time scale processes of the surface. Furthermore, one must average over many trajectories in order to acquire reasonable statistics of the hopping process and properties of interest such as the hopping distribution. All of these considerations lead to state-of-the-art simulations of surface diffusion in which thousands (or more) surface atoms are subdivided into varying hierarchies of dynamical treatments, while nevertheless requiring substantial computing power in order to integrate several thousand trajectories each on the order of a nanosecond. Unfortunately most molecular dynamics packages are not designed for surface dynamics and therefore lack the ability or needed generality to correctly and accurately construct these hierarchies, and this leads many researchers to write their own simulation codes.

Beyond the dilute adsorbate limit that is the focus of the work described here, the role of step edges, coverage, sticking and scattering in adatom diffusion has been explored by a cornucopia of computational approaches ranging from coarse-grained to ab initio methods [7, 8, 9, 55, 10, 11]. For example, it might seem that surface coverage effects can be completely described by a simple extension of the theory for single adatom hopping. However nothing could be farther from the truth as the diffusion now often includes concerted motions and may proceed by completely different mechanisms resulting in large qualitative changes in the diffusion rates [3, 29, 4, 5]. In such cases, the degree to which long-time correlations persist plays an important role in determining diffusion rates. Similar complications arise in the case of gas-surface scattering. Of particular note, within the context of reduced-dimensional descriptions, the "washboard model" —wherein the atomistic surface is replaced by an effective washboard potential— has been shown to qualitatively capture the energetic and orientational dependencies of the process [56, 57]. In cases with low energy gas-surface collisions, the sticking probability increases thereby leaving behind a transient ballistic adsorbate species that can not be described by the standard equilibrium diffusion description [3]. A more extensive discussion of the role of these features and many other relevant factors will not be discussed further. Indeed, the main focus of the work described here is on those regimes where the surface diffusion can be adequately characterized by a reduced-dimensional model in which the primary modes that are explicitly retained are sufficient to describe the adsorbate motion parallel to the surface.

1.2 Molecular Dynamics Simulations and Projective Models

1.2.1 Atomistic Models

Despite the difficulties facing computational efforts, their use has been essential in unraveling the complicated dynamics associated with surface diffusion and in determining the fundamental dissipation pathways. The seemingly simple transition from one adsorption site to its nearest neighbor is deceptively complex and is governed by a number of competing elementary mechanisms. As has been emphasized in the previous sections, the many-body coupling between the surface and adsorbate dominates surface diffusion and it governs when an adsorbate will hop and when it will stop. The primary coupling pathways can be represented by mechanical dephasings resulting from the classical interactions with the phonon bath of the surface and the quantum interactions with electron-hole pair excitations in the electronic continuum of the surface [58, 20, 21, 10, 22, 41, 3]. Generally the ratio of the vibrational frequencies of the adsorbate to that of the substrate dictates whether the phonon or electronic pathway will dominate with the latter being favored when the ratio is far from unity, although in general both mechanisms will be active to some extent [5, 10]. Within a classical framework, the bulk can be represented within a simulation as an external harmonic bath —whose spectrum is equivalent to the phonon spectrum of the underlying bulk surface up to some desired level of accuracy— to which the surface atoms are coupled [33, 34]. However an accurate theory is needed to account for the coupling of the surface to the bulk metal's electronic states. The inherent nature of the process brings into question the validity of the Born-Oppenheimer approximation given that the adsorbate does not interact with a single adiabatic potential energy surface, but rather a continuum of delocalized states associated with the metal, and relatively small changes in energy can result in excitations of electron-hole pairs [22]. As a result, Carr-Parinello dynamics involving electronic processes occurring on surfaces may not be applicable in general. Provided the coupling is weak, one possible approach that has been proposed is the "molecular dynamics with electronic friction" method designed to take electron-hole pair contributions into account resulting in a Langevin-type equation of motion for each explicit degree of freedom [20]. That is, an additional Langevin bath is introduced to describe the electronic response to adsorbate motion, albeit with its own characteristic electronic spectrum and perhaps also an effective electronic temperature that is not equal to the bulk temperature.

Several other theoretical and computational studies have been carried out that focus on particular aspects of surface diffusion and demonstrate that many factors can be important under certain circumstances. The role of lattice vibrations in the hopping process has been explored in detail by Rahman and co-workers [59, 60, 61, 32]. They note that although the rate is largely determined by the activation energy, which can in turn be accurately obtained from static high level ab initio calculations, the vibrational (entropic) contributions to the prefactor on flat surfaces and along step edges may be large and should not be ignored. In addition, a description of the diffusion of light atoms across metal surfaces must incorporate quantum mechanical effects. Hydrogen migration on a variety of substrates has been explored within the context of quantum transition state theory in which it was found that the motion of the hydrogen normal to the surface results in a significant number of barrier recrossings leading to substantial errors in the transition state theory approximation [62, 63, 64, 65, 66]. Several computational studies aimed at elucidating the effect of surface geometry on the hopping mechanisms and the ability of simulations to reproduce experimental diffusion coefficient values have been explored in detail [67, 68, 69, 70, 71]. Systematic studies of the role of coverage and long jumps have shown that both can have a dramatic impact on bulk diffusion rates [72, 73, 43, 74, 64]. These and other results have provided the foundation for general scaling arguments for the magnitude of the self diffusion constant based on macroscopic properties of the solid [31, 30].

A fundamental requirement on all of the computational studies on metal surface dynamics is the need to perform simulations with realistic potentials and in a feasible amount of time. To this end, the temperature-accelerated dynamics method [75, 14, 76] has arisen as a possible approach for reaching the latter limit. With the exception of quantum simulations, most classical simulations are based on semiempirical potentials derived either from the embedded atom method or effective medium theory [77, 78, 79]. However a recent potential energy surface for hydrogen on Cu(1 1 0) based on density functional theory calculations produced qualitatively different results from those of the embedded atom method including predictions of different preferred binding sites [80].

This brief review is in no way a complete list of all the simulations of surface diffusion that have been performed to date, but rather serves to illustrate the range of simulations that can typically be performed at present and the physical properties they can address, as well as an indication of the challenges facing current and future efforts. Moreover, it should be clear that there are substantial advantages to developing and implementing new algorithms in which one can reduce the number of atoms that must be treated explicitly within these models while retaining accuracy and the ability to calculate the requisite observables. This is precisely where projective methods can play a role, though they do not always retain the latter two requirements.

1.2.2 Projective Models

As has been described, classical molecular dynamics simulations provide a detailed description of surface diffusion but are remiss of all important quantum effects and sit on the cusp of computationally feasible simulations. As a result, many techniques for constructing coarse-grained simulations have been developed ranging from purely phenomenological descriptions to more rigorous approaches. The crux of these models lies in constructing projection operators to remove the fast degrees of freedom that are irrelevant to the process of interest [81, 82, 83]. In this manner, many modes of the surface are reduced to one effective mode influencing the surface diffusion. Usually this is described within a generalized Langevin framework for each degree of freedom of the adsorbate whose equation of motion is of the form

$$m\dot{v} = -m \int_0^t \gamma(t-t')v(t')\,dt' + \xi(t) - \nabla U(q) \;, \tag{3}$$

where $\gamma(t-t')$ is a time-dependent friction term including the entire history of the trajectory, $\xi(t)$ is a zero-mean stochastic process and $U(q)$ is the deterministic potential of mean force experienced by the adsorbate. This equation of motion is equivalent to a Hamiltonian system in which the primary degrees of freedom are coupled bilinearly to a bath of harmonic oscillators [82, 83, 84]. The origin of the stochastic force in the Langevin equation is related to the undetermined initial conditions of the bath in the Hamiltonian system. The former is connected to the friction by a fluctuation-dissipation relation that ensures the system approaches thermal equilibrium in the steady state [85]. Equation 3 is in no way more tractable than the Hamiltonian approach until one assumes a form for the friction kernel that is generally short-lived and physically corresponds to a quick response of the bath to the system motion. Calculations have shown that this approximation is usually justified particularly in the context of surface diffusion [37, 38, 3, 86]. Following this prescription, Tsekov and Ruckenstein have developed an exact Langevin description of the surface within the harmonic approximation for the coupling of the surface atoms [35, 36]. Additionally in a series of papers, Tully and co-workers have developed a model in which a small number of surface atoms closest to the adsorbate are treated explicitly [52, 53, 33, 34, 54]. These are then coupled to a Langevin bath constructed to represent the bulk phonon spectrum of the surface.

These studies have paved the way for more phenomenological approaches that although less rigorous have nonetheless provided useful insights into the diffusion process. It is generally accepted that the potential of mean force $U(q)$ on which the adsorbate moves is corrugated and can be adequately approximated by a periodic function whose period is related to the lattice spacing of the surface [38, 28, 3]. The barrier heights impeding diffusion are derived from the activation energy determined either experimentally or from a higher level of theory. Furthermore, the friction kernel typically decays on a very short time scale on the order of a few hundred femtoseconds [38] and the magnitude lies in the low to intermediate friction regime [37, 28]. The fast decay is much shorter than any time scale related to surface diffusion and so the generalized Langevin equation can be approximated by the simplified memoryless Langevin equation

$$m\dot{v} = -m\gamma v + \xi(t) - \nabla U(q) \ . \tag{4}$$

This adiabatic approximation assumes the bath is capable to instantaneously responding to the motions of the subsystem.

At this level of description, the simulations are so inexpensive that it is possible to explore the full range of the parameters and obtain converged results. Often here one is not concerned with making quantitative comparisons with experimental values but rather in extracting general qualitative trends or in searching for new phenomenon [87, 88, 89, 90]. Using such models, several groups have employed phenomenological Langevin equations to probe the role of friction and how the particular structure of the potential of mean force can influence surface diffusion [91, 92, 93, 94, 95, 96]. Alternatively, as for example in references [28, 29], through comparison with experimental data the authors were able to deduce the form of the potential and friction so as to reproduce experimental results providing valuable insights on the nature of the system. Within this formalism, the subject of multiple hops has received considerable attention. It has been shown that the probability of observing a long jump can be strongly influenced by both the form of the potential as well as the friction [91, 93, 87, 97, 98, 99, 39, 37, 40, 100, 95, 96]. The ability of these models to probe a large range of parameters provides bounds for which to look for new phenomenon experimentally or at a higher level of simulation, as well as an intuitive and simplified picture of the process which can often be obscured by the sheer amount of data produced by all-atom simulations.

As has been emphasized throughout this introduction there are many length and time scales that influence adsorbate diffusion from couplings with electron-hole pairs to surface phonon modes all of which have been shown to be non-negligible in certain instances [101, 58, 10, 22, 32, 60]. However, in most coarse-graining procedures all of these effects are averaged into a single effective mode governed by a simple (generalized) Langevin equation with a static potential of mean force. A central question remains as to whether this is an adequate description. If not, how should one go about constructing a reduced dimensional model that is capable of retaining some of these features. Keeping with the simplicity of the Langevin framework, there are only two possibilities to account for the multi-scale factors, the potential of mean force or the friction term and noise terms. Recent advances involving the coupling of a particle to multiple or time-dependent baths leading to modifications of the friction kernel and the noise term is one possibility that may provide some insight to the problem [102, 103]. Alternatively one could incorporate the multi-scale details in the potential of mean force by allowing for time-dependent variations in the structure as have been explored by several authors [104, 105, 106, 107, 108, 109]. The latter approach is one of the main features described in the following chapters.

1.3 Overview

Chapter 2 describes molecular dynamics simulations of adsorbate diffusion on metal surfaces in the spirit of those described earlier in Section 1.2.1. The shear computational effort required for this study necessitates the development of the coarse-grained simulations discussed in the remaining chapters. In particular, calculations in the low temperature regime become increasingly impractical due to the poor hopping statistics. However, this fact does not undermine the utility of such simulations and at a larger temperatures, a detailed atomistic description of the diffusion path and mechanism can be obtained. In this chapter, parallel classical simulations are constructed and carried out to provide estimates for the various parameters required for Langevin simulations. Calculations of the diffusion coefficient are first performed to validate the method, and then the dynamic friction and potential of mean force are extracted. These simulations serve as the groundwork for the work carried out in Chapters 3 and 4.

Chapter 3 provides a more thorough introduction to the Langevin dynamics described in Section 1.2.2. The primary extension discussed herein is the incorporation of the multiple time scale interactions mentioned at the end of the preceding section through a time-dependent stochastic potential of mean force. A standard feature associated with models of diffusion over a single stochastic barrier is the so-called “resonant activation” [104] in which the rate of passage over the barrier may be strongly affected by the associated fluctuations in the underlying potential. The key questions addressed in this work are whether this phenomenon will continue to be observed in periodic or quasi-periodic one- and two-dimensional systems resembling those that are often used in the study of surface diffusion, and if so how robust this phenomenon is with regards to the particular nature of the potential. It is found that the resonant activation does indeed exist in extended systems and that it is largely insensitive to the details of the potential of mean force. Therefore, it is proposed that the ability to spatially control this phenomenon would provide one with the ability to direct diffusive motion along the surface [90].

Chapter 4 continues along the lines of Chapter 3 focusing on an alternative method to include the dissipative term in the Langevin equation [110]. In the presence of a stationary potential of mean force, the damping arises only from the fluctuations in the surrounding bath, *viz.* the noise. However, if the potential is also stochastic then an additional frictional term must be incorporated to ensure the system remains in thermal equilibrium. In previous work [109, 111, 90], this has been achieved through an iterative self-consistent procedure such that this requirement is strictly enforced. In this chapter, an alternative method based on an approximate analytic form is derived and the limits of its validity are addressed. In the wider context, this analysis demonstrates that stochastic potentials provide an alternative origin of the space-dependent friction that has been the focus of much attention in the literature.

The remaining chapters detail fruitful collaborative work on topics largely unrelated to surface diffusion. Chapter 5 discusses a novel approach to transition state theory (TST) in noisy environments. The reaction rate is a key factor in many chemical systems and reasonable estimates for this quantity can generally be obtained from transition state theory or one of its descendants. In isolated systems such as gas phase chemical reactions, TST may indeed provide an accurate approximation, but unfortunately many interesting chemical systems are noisy (i.e. governed by a stochastic differential equation typified the Langevin equation) due to complex interactions with the surrounding environment. Consequently in these cases TST most always provides only an upper bound to the true rate. While many previous theoretical treatments of this problem have taken advantage of the equivalence of the Langevin equation with the corresponding Hamiltonian or Fokker-Planck description in which the noise in included implicitly, here we stay within the Langevin framework. The elegant approach of Bartsch et al. [112, 113] based on the construction of a special reference trajectory that gives rise to a noiseless coordinate system is applied in numerical simulations of prototypical reactive systems in order to address the utility of this method with respect to traditional approaches [114].

Finally, Chapter 6 outlines the construction of a new checking function that may prove useful for protein structure validation [115], and its ensuing incorporation into a web server to facilitate its use in the community [116]. One of the grand challenges facing theoretical and computational scientists is to develop methods capable of reconstructing the complicated secondary and tertiary structure of a protein given only sequence information. Not surprisingly, current efforts take advantage of several well-established metrics to judge the quality and consistency of a given proposed structure. This chapter provides an additional check based on dihedral angle analysis to aid in protein modeling. With the ever-increasing number of protein structures determined by experimental or computational methods, the existing database of information contained therein grows accordingly giving rise to a statistical ensemble that can be readily probed to provide statements on the compatibility of a given structure with the existing database. In Chapter 6, this approach is employed to extend the standard analysis of dihedral angle correlations within a given amino acid to correlations that may exist between neighboring residues. Based on these results, a new checking function for protein structure analysis is introduced that is rooted in information theory and has subsequently been incorporated into a convenient web-based tool for use in the community.[1]

---

[1] Aside from Chapter 2, this introduction and the remaining chapters largely represent material that has been published or submitted as an original scientific work [90, 110, 115, 114, 116, 117].

CHAPTER II

REDUCED DIMENSIONAL MODELS DERIVED FROM MOLECULAR DYNAMICS SIMULATIONS OF TRANSPORT ON METAL SURFACES

2.1 Introduction

Diffusion on metal surfaces remains an area of intense study due to the intriguing physics governing the fundamental processes involved as well as its implications in a wide range of technological applications [12, 11, 10, 5]. However, as was discussed extensively in the preceding chapter, the large dimensionality of the system and the multiple time scales involved are the source of substantial difficulties for both experiment and theory [24, 25, 26, 118, 119]. Furthermore, the complexity of a model that is capable of incorporating all of these dependencies may fail to shed light on the elementary mechanisms involved and can often obscure simple, physical interpretations. In these situations, it would be advantageous to develop reduced descriptions of the process that are capable of capturing the essential features of the diffusive motion of the adsorbates [34, 35, 36, 38, 12, 11]. Such models provide an intuitive description of the system by projecting out all of the non-essential degrees of freedom and retaining only those that are vital to the process at hand. From a computational perspective, these models are ideal since they are exceedingly efficient while still capable of yielding reliable results. Unfortunately, such models may sometimes be too coarse or lack the requisite transferability to make their development worthwhile. For example, in some instances the full-dimensional description is absolutely necessary to analyze microscopic level details that are not present in the coarse-grained description. Obviously, the optimal situation is to have access to both the complete atomistic description of the system as well as its coarse-grained counterpart and to have the ability to readily convert between the two based on the length and time scales of interest. Of course, the viability of such an approach is contingent upon the fact that both models are correctly describing the physical process of interest and are capable to providing meaningful results and predictions.

The purpose of this chapter is to create such a model and demonstrate that the reduced-dimensional description is capable of providing results in agreement with higher level atomistic simulations. A model potential that is capable of reproducing the qualitative trends observed experimentally for self diffusion on the (1 0 0) surface of nine different metals is used in full molecular dynamics simulations. The diffusion coefficients are extracted and a comparison between the current results and those obtained in previous studies is presented. Then for the silver surface, a coarse graining procedure is applied in which every atom of the surface is projected out leading to a Langevin model of adsorbate motion on an effective potential of mean force. However, the method is completely general and readily allows one to retain some number of surface atoms if they are required for the process at hand as for example in gas-surface collisions or when exploring the site exchange mechanism [50, 51]. In many situations there is no reason to believe that the Langevin model which is equivalent to a subsystem coupled to a harmonic bath should provide an accurate description of the system. However, for surface diffusion this approximation should be rather good since the substrate may often be quite reasonably described within a harmonic representation. Within this framework, the potential of mean force and the dynamic friction required for the Langevin description are extracted. Approximations for these values based on purely physical considerations have served as the starting point for many studies of surface diffusion [52, 53, 33, 34, 54]. An additional outcome of this work is that the simple phenomenological models of this type are shown to be rather reasonable. Finally an analysis is presented of the reduced-dimensional system and how estimates of the rate can be computed from transition state theory without recourse to any further numerical simulations. This result is of particular interest in the low-temperature regime where brute force dynamics simulations become computationally impractical.

2.2 Molecular Dynamics Simulations

Classical molecular dynamics simulations of self diffusion on the (1 0 0) surface of Ag, Al, Au, Cu, Ir, Ni, Pd, Pt and Rh have been performed. This is an analogous set to the systems studied by Feibelman [31] with density functional calculations and a subset of those examined by Agrawal et al. [30] with classical Monte Carlo techniques. Remarkably, these two very different simulations arrived at the same conclusion that the self diffusion coefficient for a variety of metals can be reliably estimated from macroscopic properties of the surface. This result is also reproduced here with molecular dynamics simulations.

In this work, the metal surface is represented as a mixed-ensemble perfect slab that is ten layers thick each of which contains 400 atoms. Periodic boundary conditions are applied in the plane only. The bottom two layers are held fixed with positions given by the experimental lattice spacing [120]. The next four layers are kept at constant temperature using a chain of Nose-Hoover thermostats [121, 122]. The remaining layers and all adsorbates are simulated with undissipated Newtonian dynamics using a time step of 1 fs [34, 123]. Although the top layers are not dissipated the coupling to the lower thermostatted layers leads to simulations that are not strictly energy conserving. The advantage of this approach is that the surface can be held at a constant temperature (after some equilibration) and therefore the diffusion coefficient is well defined so that an activation energy and prefactor may be extracted. However, the dynamics of the adsorbate and its nearest neighbors are largely unaffected by the damping introduced by the thermostats. In the simulations described here, the particular number of NVT layers was observed to have a negligible impact on the average properties, although all were performed at rather high temperatures. The surface atoms are modeled by a simple potential consisting of only Van der Waals interactions with the parameters that were developed in reference [30]. Although admittedly this description is not completely adequate, it is sufficient to demonstrate the main purpose of this work that coarse-grained models can capture the essential features of the corresponding higher level model. For computational efficiency, the pairwise interactions for each atom in the bulk are evaluated up to third-nearest neighbors. A larger neighbor list size does not significantly change the results. The forces on the adsorbates, however, include a much larger neighbor list. Upon initialization, the system is equilibrated for 1 ns. The subsequent 49 trajectories are run on the same thermalized slab following an appropriate randomization of the adsorbates initial conditions and a brief 10 ps equilibration.

2.2.1 Diffusion Rates

The diffusion coefficient has been determined from the standard relations for the mean-squared displacement

$$4Dt = \langle (r(t) - r(0))^2 \rangle \ , \tag{5}$$

and the velocity autocorrelation function

$$D = \frac{1}{2} \int_0^\infty \langle v(0)v(t) \rangle dt \ . \tag{6}$$

In the case of the symmetric (1 0 0) surfaces under study, only the isotropic diffusion in the plane of the surface is considered, and each adsorbate is left unconstrained with respect to the coordinate normal to the surface. Following the equilibration periods described above, the relevant quantities in equations (5) and (6) for the adatom are calculated from 1500 independent trajectories of 135 ps each. In the standard manner, the slope of the mean squared displacement or the integral of the velocity autocorrelation function of the diffusing atom yield the diffusion coefficient. These two approaches produce the same result within statistical error. As is well known, the hopping is an activated process and as such the diffusion coefficients for each metal obey an Arrhenius-like equation,

$$D = D_0 \exp\left(-E_\mathrm{a}/k_\mathrm{B}T\right) \ , \tag{7}$$

allowing one to extract the activation energy and prefactor from a collection of diffusion coefficients measured over a series of temperatures. The diffusion profiles for the nine metals are displayed in Figure 2 along with the results reported in reference [30]. Rather good agreement is observed between the current simulations and the previous study even though the latter were calculated by Monte Carlo techniques (cf. equation 2). Moreover, the results scaled to the appropriate dimensionless units, $T_0 = k_\mathrm{B}T/\epsilon$ and $D_0 = D\sqrt{m/\epsilon\sigma^2}$ where ϵ and σ, and m represent the respective potential parameters and mass for a given

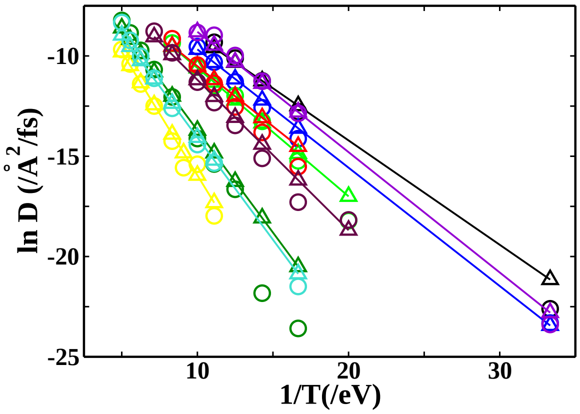

Figure 2: Diffusion coefficients depicting the results of the present MD simulations (circles) along with those derived from the prefactors and activation energies reported previously in reference [30] displayed as lines and triangles. The black, blue, green, red, violet, yellow, maroon, dark green, and turquoise lines reflect the diffusion on silver, copper, gold, palladium, aluminum, iridium, nickel, rhodium and platinum surfaces, respectively. Each point was obtained from the slope of the adatom's mean-squared displacement averaged over 1500 independent 135 ps trajectories.

metal, lead to universal behavior for all metals as seen in Figure 3 [31, 30]. However, at low temperatures, the statistical error in the current method becomes obvious.

2.3 Reduced-Dimensional Systems

In coarse-grained simulations, the potential of mean force (PMF) plays a central role in governing the dynamics. Calculations have been performed to extract the PMF experienced by an adatom diffusing across the silver surface. The procedure involves calculating the forces on a ghost atom placed at given positions r_a above the surface,

$$\frac{dw}{dr_a} = \left\langle \frac{\partial U}{\partial r_a} \right\rangle , \tag{8}$$

where the average is taken over different configurations of the substrate. Once obtained, the PMF $w(r)$, is obtained by integrating the mean forces from the full dimensional system. Many positions r_a must be evaluated in order to achieve a sufficient representation of the underlying surface. Fortunately, this challenging procedure can be simplified by taking advantage of the high symmetry of the surface. One needs to only evaluate the forces at specified positions inside of a unit cell enabling one to average over every cell on the surface. In practice, 500 independent trajectories are simulated for 65 ps each with the

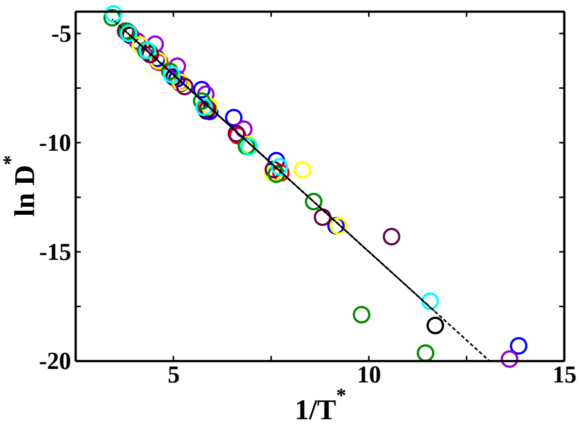

Figure 3: Diffusion coefficients for each of the metals displayed in Figure 2 in appropriate dimensionless units. The black, blue, green, red, violet, yellow, maroon, dark green, and turquoise lines reflect the diffusion on silver, copper, gold, palladium, aluminum, iridium, nickel, rhodium and platinum surfaces, respectively.

forces evaluated on a three-dimensional grid containing 1000 points above each unit cell of the surface. Once obtained, the ensuing integration of equation (8) is performed. Since the PMF is a state function, statistical error can be minimized by averaging the results of integrating many paths between points on the grid of average forces. Between nearest neighbors approximately 50 independent random walker paths are integrated and the results averaged to produce $w(r_a)$. In this way, the PMF evaluated at a constant 1.85 Å (the approximate spacing between layers in the lattice) above the silver surface unit cell is displayed in Figure 4 and the one-dimensional projection along the reaction coordinate (the diagonal component of Figure 4) is displayed in Figure 5.

The friction can most readily be obtained by evaluating the fluctuation-dissipation relation

$$\gamma(t-t') = \frac{1}{mk_{\mathrm{B}}T}\langle \delta F(t')\delta F(t)\rangle \ , \tag{9}$$

and realizing that the noise in the Langevin description simply represents the deviations of the instantaneous force from the average, $\delta F(t) = F(t) - \langle F \rangle$ in the molecular system. It is efficient to simultaneously evaluate the requisite force autocorrelation function at each point in the grid during the course of the potential of mean force calculation. As displayed in Figure 6, the friction consistently displays a very short decay time on the order of 200

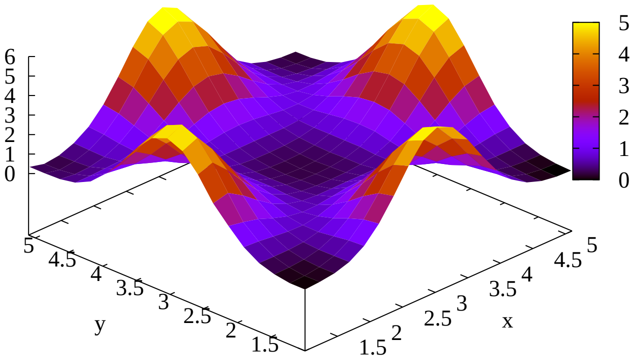

Figure 4: The potential of mean force evaluated at a fixed height 1.85 Å above the surface. The peaks correspond to equilibrium positions of atoms in the top layer of the surface while the minima correlate with the positions of atoms in the first sublayer. Diffusing adsorbates traverse the relative maxima between the two.

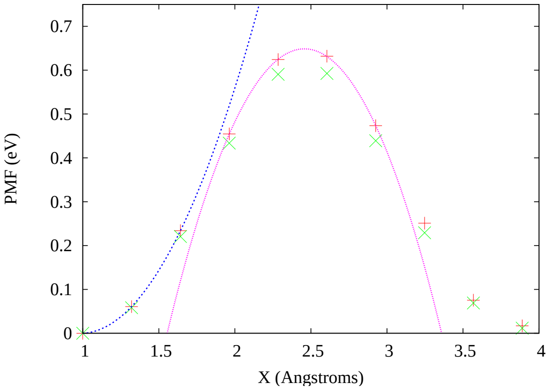

Figure 5: The potential of mean force (red crosses) along the reaction coordinate evaluated at a fixed height 1.85 Å above the surface. The green points display the average potential and the solid lines represent harmonic fits near the well bottom and barrier providing a frequency of $\omega = 6$ ps^{-1} and barrier height 0.648 eV.

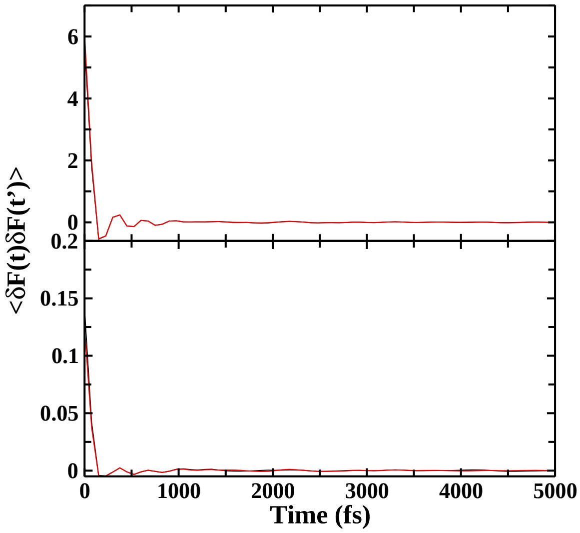

Figure 6: The force autocorrelation function $(\mathrm{eV}/\text{Å})^2$ evaluated at the well (lower) and barrier (upper) on the Ag surface at a temperature of 0.08 eV and a fixed height 1.85 Å above the surface. Within the Ohmic approximation, this corresponds to the very low friction regime (0.3/fs at the barrier and 0.01/fs at the well).

fs regardless of whether it is evaluated at a minimum or a maximum along the reaction coordinate. This fast relaxation time was also observed in reference [38]. Since all motion relevant to surface diffusion occurs on a much longer time scale, this observation further justifies the standard use of Ohmic friction in Langevin simulations of surface diffusion. The magnitude is determined by replacing the dynamic friction with a delta function of the same area. These simulations result in a value typically lies in the low friction regime ($\gamma \ll \omega$), although the difference in magnitude between the well and the barrier is more than an order of magnitude possibly leading to the need for a space dependent friction.

2.4 Conclusions

In this chapter, a rigorous approach to obtaining the relevant parameters required for a coarse-grained description of surface diffusion from full molecular dynamics simulations has been performed. Within the Langevin representation, an overall reduction of the computational effort is achieved that scales with the number of atoms in the slab. This fact becomes all the more advantageous as the size and complexity of the adsorbed species increases. Furthermore, with these quantities in hand, it becomes trivial to evaluate many important properties associated with surface diffusion. From the nature of the PMF the location of the preferred adsorption site is immediately obvious. Likewise, by combining the knowledge of the activation energy obtained from the PMF with the value of the friction constant, an accurate estimate of the rate may be obtained directly from transition state theory [124, 125, 47]. The coarse-grained description developed here and employed in the next two chapters allows one to circumvent the costly atomistic dynamics when searching for new and interesting phenomenon while retaining an accurate representation of the system.

CHAPTER III

A PHENOMENOLOGICAL MODEL FOR SURFACE DIFFUSION: DIFFUSIVE DYNAMICS ACROSS STOCHASTIC POTENTIALS

3.1 Introduction

As was demonstrated in the preceding chapters, theoretical efforts are capable of qualitatively reproducing experimental results by employing simple phenomenological models consisting of a periodic substrate resembling the atomic unit cell with barrier heights to diffusion on the order of the activation energy [33, 34, 126]. As an extension to these models for the surface, we have previously utilized a periodic, stochastic potential of mean force such that the barriers are no longer static in time [111]. This inclusion enables the equation of motion to account for both the electronic structure and time-dependent nature of the surface provided that its statistical properties are known. Additionally it was essential to extend this description into the low-friction regime in order to accurately account for the weak interactions of the adsorbate with the surface as has been observed by several experiments [127, 128] and the results of the previous chapter. Perhaps not surpringly, some peculiarities have been noted with this class of stochastic potentials [129, 104, 130]. In particular, the rate of diffusion across the surface is heavily dependent upon the fluctuations of the barriers leading to a maximum in the rate of transport when the inherent time scale of the dynamics is of the same order as these fluctuations. This phenomenon, which was first termed "stochastic resonance" in a model for climatic changes [129] when the driving potential is deterministic (and likely periodic), and later [104] "resonant activation" when the potential is nondetermistic has been the subject of considerable research over the past decade, and is the primary motivation for this chapter [129, 104, 130, 108, 39, 97, 89, 131].

The ultimate goal herein lies in the possibility of providing a novel means for patterning a surface at the atomic level. Theoretically, it has been shown that the stochastic resonance persists in two-dimensions when the process is driven by an external field [88]. Additionally, we have shown that if one can instead introduce internal fluctuations into a one-dimensional surface, then it is possible to enhance the diffusion of adsorbates [111]. Experiments have proven that electron and photon bombardment of surfaces can enhance adsorbate diffusion [132, 133]. It has also been shown recently through a series of experiments that the electric field produced by a scanning tunneling microscope can stimulate adatom diffusion toward the area around the tip on a room temperature metal surface, and that this can be done in a non-uniform manner leading to the formation of complex structures at the sub-nanometer scale [134, 135, 136]. Here, we propose a more general mechanism based upon resonant activation that can be used to control the diffusive behavior in a less intensive manner by stimulating the surface on a broad scale. If one could further obtain some spatial control over these fluctuations, then it is conceivable that one could pattern a surface in any way desired in a non-destructive manner using the intrinsic properties of the surface to guide the path of the adsorbate.

In the present case, we have extended the phenomenological equation of motion to include two-dimensional, stochastic potentials as a first step toward a more realistic model of the adatom dynamics, as well as to provide fundamental insights into the mechanisms governing these systems. Will nontrivial effects arise due to coupling between the spatial degrees of freedom? Recent evidence from several experiments have concluded that diffusion processes involve multiple, complex hopping mechanisms to both nearest and non-nearest neighbors which suggests that it may be possible [127, 128, 118]. For example, it has been proposed that an adatom that originally traverses linearly along one direction may later change its course during its interactions with the surface and diffuse along the other [118]. In this case a one-dimensional potential will never be capable of modeling the dynamics properly. The results shown here are in agreement with this picture in that the two-dimensional transport properties are found to be non-trivially related to the corresponding one-dimensional results.

In Section 3.2, the heuristic construction of the equation of motion for an underdamped particle experiencing a stochastic potential is presented, as well as a brief description of the characterization of the hopping process. Section 3.3 presents a comparison of the dynamics seen on two different variants of the two-dimensional surface and their corresponding one-dimensional counterparts. As in earlier work, the numerical rates obtained for a surface at the limits of the correlation time in the stochastic potential are computed as a verification for the intermediate regimes.

3.2 Model and Methods

The diffusive behavior of a Brownian particle traversing a stochastic potential is governed by the Langevin equation (LE)

$$\dot{v} = -\gamma v + \xi(t) + F(q;t) \ , \tag{10}$$

where $v = \dot{q}$ is the velocity of the particle, γ is the friction constant, and $F(q;t)$ is a stochastic external force. For clarity, the mass of the particle has been set to unity through a canonical transformation, but is readily included. The thermal fluctuations present in the system are described by $\xi(t)$, and are treated in the standard manner as a Gaussian white noise source with zero mean and correlation determined by the fluctuation-dissipation theorem:

$$\langle \xi(t)\xi(t')\rangle = 2k_{\mathrm{B}}T\gamma_{\mathrm{th}}\delta(t-t') \ . \tag{11}$$

In the presence of a stochastic potential, the friction constant must be renormalized to balance this additional noise in order to satisfy equipartition, especially in the low friction regime where these contributions may become significant. This has presently been accomplished through a self-consistent procedure described previously, [109] in which γ is renormalized according to the relation

$$\gamma^{(n+1)} = \gamma^{(n)}\left(\frac{\langle v^2(t)\rangle_n}{k_{\mathrm{B}}T}\right) \ , \tag{12}$$

at each n-th iteration until convergence. An alternative that includes the spatial dependence of the friction has been derived based upon an extension of the fluctuation-dissipation theorem, and is main focus of the following chapter. In brief, the major criticism to the current approach lies in the approximation made in developing equation (12) in which $F(q;t)$ is treated as a local noise source obeying a fluctuation-dissipation relation equivalent to equation (11). However, the stochastic potentials discussed here have memory and are therefore nonlocal in nature leading to non-vanishing cumulants at third and higher orders. These effects are included, but only in an average manner, in the self-consistent approach and are ignored in the space-dependent approach all together. However, at this level of description, both methods are capable of capturing the essential dynamical effects as will be demonstrated in Chapter 4 Here, for convenience we will employ the simpler self-consistent approach.

The stochastic potential gives rise to a time-dependent force, $F(t) \equiv -\nabla_q U(q;t)$. In the model studies that follow, the form of the potential, $U(q;t)$, is either that of a set of merged harmonic oscillators (MHOs) or sinusoidal. The MHO potentials can be represented explicitly as

$$U(x;t) = \begin{cases} \frac{1}{2}k_0(x - x_m^0)^2 & \text{for } x_m^0 < x \leq x_m^- \\ V_m^\ddagger + \frac{1}{2}k_m^\ddagger(x - x_m^\ddagger)^2 & \text{for } x_m^- < x \leq x_m^+ \\ \frac{1}{2}k_0(x - x_{m+1}^0)^2 & \text{for } x_m^+ < x \leq x_{m+1}^0 \end{cases} , \tag{13}$$

where the wells and barriers are centered at $x_m^0 = -\lambda/2 + m\lambda$ and $x_m^\ddagger = m\lambda$, respectively. The connection points are constructed to ensure continuity in the potential and its first derivative; this leads to the result $x_m^\pm = \pm k_0\lambda/(2k_0 - 2k_m^\ddagger) + m\lambda$. The width of the barriers vary stochastically in time according to the relation $k_m^\ddagger = -(k_0 + \eta(m,t))$. This, in turn, defines the barrier height $V_m^\ddagger = -k_0 k_m^\ddagger \lambda^2/(8k_0 - 8k_m^\ddagger)$.

In this work, the two-dimensional sinusoidal potentials are given the simple form,

$$U(x,y;t) = (2E_\mathrm{b} + \eta(t)) \left(\sin\left(\frac{2\pi x}{a}\right) \sin\left(\frac{2\pi y}{a}\right) + 1 \right) , \tag{14}$$

where the wells are connected by barriers along the diagonals, and the one-dimensional potentials are similarly defined as

$$U(x;t) = \left(E_\mathrm{b} + \frac{1}{2}\eta(t)\right) \left(\sin\left(\frac{2\pi x}{b}\right) + 1 \right) . \tag{15}$$

The two have equivalent well-to-well distances if $b = a\sqrt{2}$. (See Appendix A.0.1 for the connection between equation (14) and the rectilinear form chosen in reference [91].) In both cases, E_b is the barrier height and $\eta(t)$ is an auxiliary stochastic variable. In the numerical investigations, the thermal energy is chosen so that it is 1/6 of the average value of the barrier heights. The constant factors in the amplitudes are included to insure consistency between the barrier heights in the two models. In two-dimensions, particles selectively escape over the saddles of the potential, as opposed to the one-dimensional case in which they are forced to traverse the maxima. As such, one can reference the values of the instantaneous barrier height for both the one- and two-dimensional potentials to the chosen temperature such that it has the values, $6k_{\mathrm{b}}T + \eta(t)$.

As can be seen from the form of the equations above, these potentials fall into two general classes based upon the spatial correlations of the barrier heights. For instance, all barriers of the sinusoidal potential fluctuate in unison, $i.e.$ coherently, with one another, while the barriers of the MHOs are capable of fluctuating independently, $i.e.$ incoherently. Nonetheless, in the limit that the MHOs are forced to fluctuate coherently, the transport properties that it gives rise to are nearly identical to those found for the corresponding stochastic sinusoidal potential. Consequently the numerical results for the one-dimensional coherent stochastic MHO potential have been omitted throughout this chapter. For simplicity, the one-dimensional stochastic *coherent* sinusoidal potential and *incoherent* MHO potential will be referred to as $1\mathrm{D}_{\mathrm{sin}}$ and $1\mathrm{D}_{\mathrm{MHO}}$, respectively. In addition to the two-dimensional coherent sinusoidal potential, a quasi-incoherent two-dimensional sinusoidal potential has also been included by allowing the barriers in the two respective dimensions to fluctuate independently of one another by including a stochastic variable for each dimension. These will henceforth be referred to as the coherent, $2\mathrm{D}_{\mathrm{C}}$, and incoherent, $2\mathrm{D}_{\mathrm{I}}$, potentials. A fully incoherent two-dimensional surface has been developed in which every barrier is allowed to fluctuate independently of all other barriers, and is currently being tested [137].

As mentioned previously, the stochastic features of these potentials are contained in $\eta(t)$, which is, in turn, defined as an Ornstein-Uhlenbeck process. It is described by the well-known stochastic differential equation,

$$\dot{\eta} = -\frac{\eta}{\tau} + \sqrt{\frac{2\sigma^2}{\tau}}\zeta(t) \; , \tag{16}$$

where σ^2 is the variance of the distribution, τ is the η-correlation time in the stochastic potential, and $\zeta(t)$ is an additional white noise term. As such, $\eta(t)$ has a probability distribution,

$$P(\eta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\eta^2}{2\sigma^2}\right) , \tag{17}$$

and is correlated in time through the expression,

$$\langle \eta(t)\eta(t')\rangle = \sigma^2 \exp\left(-\frac{|t-t'|}{\tau}\right) . \tag{18}$$

Modification of τ allows one to control the extent of the exponentially decaying correlation in the barrier heights. Simulations are performed over ten decades in the η-correlation time, τ. In the lower limit, a fully random surface with variations in the barrier heights from one instant in time to the next is parameterized only by the variance. The higher limit essentially corresponds to the sampling of single realizations of the potential energy surface whose barrier heights are determined by the initial values of η as chosen from the distribution given by equation (17).

The dynamics of these systems were characterized by the mean first passage time (MFPT) of a particle to escape from an initial minima of the potential. The corresponding rate for such a process is given by the inverse of this quantity. However, one must be cautious in defining when a particle has escaped from its initial well. A standard transition state approach is not adequate when considering this class of potentials because the barrier heights (and possibly positions) are not fixed. A method capable of accounting for this type of phenomena is the use of a geometrical constraint in the phase space of the Brownian particle [138]. More precisely, the criterion used to characterize a first passage process is defined as the time required for a particle to escape its present minima and stabilize with an energy $E \leq Dk_\mathrm{b}T$ in another minima, where D is the dimensionality of the system. The MFPT is thus defined as the average of the first passage times over all trajectories,

$$\tau_\mathrm{MFPT} \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \tau_\mathrm{FPT}(i) . \tag{19}$$

Stochastic trajectories were obtained by numerical integration of the coupled equations (10) and (16) according to the method described by Ermak and Buckholtz [139, 140]. For these simulations, the MFPT was seen to converge with respect to both the time step size and number of trajectories when their values were 0.001 and 10,000, respectively.

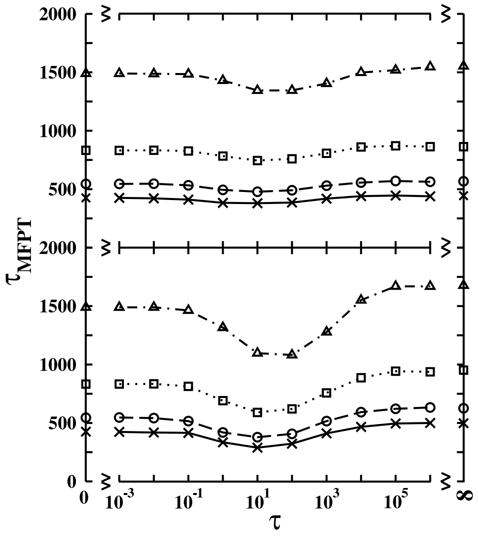

Figure 7: The MFPT versus the correlation time, τ, at $k_{\mathrm{B}}T = 2/3$ (in units of some standard temperature $k_{\mathrm{B}}T_0$) on the $2\mathrm{D_I}$ potential in the overdamped regime for values of the friction constant, γ_{th}: 0.8 (solid curve and x symbols), 2 (dashed curve and circles), 4 (dotted curve and squares), and 8 (dot-dashed curve and triangles). The variance, σ^2, is equal to 0.2 and 0.6 in the top and bottom panels, respectively. The calculations for the MFPTs in the $\tau = 0$ and $\tau = \infty$ limits are described in the text and are presently displayed on the ordinates.

3.3 Results and Discussion

As a first test of our approach, the dynamics were studied in the overdamped regime. This regime has been studied extensively over the last decade, largely due to the phenomenon known as resonant activation [104]. Depending on the η-correlation time, τ, of the fluctuations of the potential, a minimum in the MFPT (maximum in the rate) is observed for these systems. The results for our simulations in the high friction regime are displayed in Figure 7 for the $2\mathrm{D_I}$ potential, with the top and bottom panels representing two different values of the variance of the distribution given in equation (17). The correction to the friction constant due to the stochastic potentials is negligible compared with the large thermal component in this regime. The values at the ordinates represent the numerically calculated values of the MFPT for the $\tau = 0$ and $\tau = \infty$ limits. In the zero η-correlation time limit, the fluctuations in the potential are so rapid that the diffusive particle effectively experiences the average, stationary potential, from which the dynamics are computed. At the opposite limit, however, the barrier height is essentially a fixed value representing dynamics on a single realization of the surface whose magnitude is sampled from the distribution given by equation (17). As can be seen, the behavior of the MFPT calculated for intermediate values of the η-correlation time is approaching the respective limits, which lends weight to our physical understanding of the phenomenological prescription.

Until recently, these simulations have only been performed in the overdamped regime due to an inability to accurately describe the friction in the underdamped regime where additional dissipative factors must be included to maintain equipartition. Using the method described above for constructing this correction, we have extended these simulations into the low friction regime. If there is no coupling between the two degrees of freedom, $i.e.$ the passage time is simply related to the probability of escape, then the rate (inverse MFPT) on the two-dimensional surfaces should be twice that of the corresponding one-dimensional potential. Plotted in Figure 8 and 9 are the MFPTs for two values of the variance obtained for each of the three sinusoidal surfaces with those from the 1D model scaled by a factor of 1/2 to test this hypothesis. The well-to-well distance is $\sqrt{2}$ for all three potentials. The top, middle, and bottom panels correspond to values of the friction constant of 0.08, 0.2, and 0.4, respectively. For low to intermediate values of the η-correlation time, τ, the behavior of the MFPT on the one-dimensional potential essentially mirrors that of the two-dimensional coherent potential except for a constant shift to lower passage times that is invariant to changes in friction or variance. This can be attributed to topographical differences between the two surfaces. For example, in one-dimension, the particle may surmount the barrier if it acquires enough energy, regardless of its corresponding direction of travel. However, the two-dimensional potentials includes four areas of almost insurmountable energy that the particles must avoid in order the escape, thus increasing the passage time.

At low η-correlation time, the MFPT's for both of the two-dimensional potentials are coincident. This is the result one would expect since this regime corresponds to the dynamics of the average potential in which the two surfaces are equivalent. As τ is increased, and hence the surface behaves less like the average, the differences between the two-dimensional models becomes significant, especially when the variance of the distribution controlling

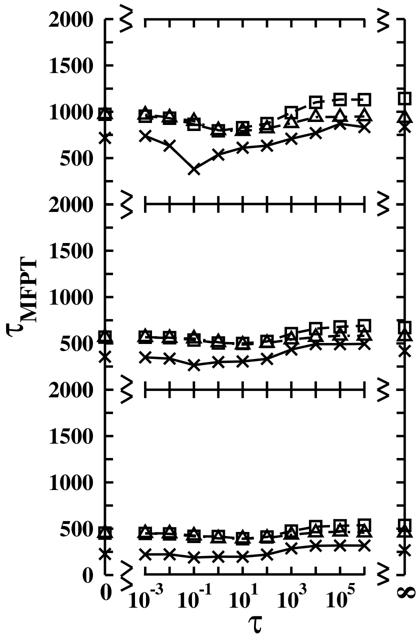

Figure 8: The MFPT versus the correlation time, τ, in the underdamped regime for each of three potentials: $1\mathrm{D}_{\mathrm{sin}}$ (solid curve and x symbols), $2\mathrm{D}_{\mathrm{C}}$ (dashed curve and squares), and $2\mathrm{D}_{\mathrm{I}}$ (dot-dashed curve and triangles). The top, middle, and bottom panels represent results obtained at the values, 0.08, 0.2, and 0.4, of the friction constant, γ_{th}, respectively. The variance, σ^2, is equal to 0.2 and the well-to-well distance is $\sqrt{2}$ in all three cases. The limits displayed on the ordinates were obtained from numerical simulations as in Figure 7.

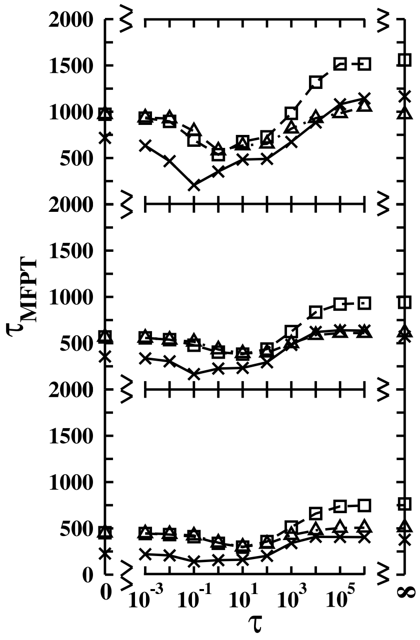

Figure 9: The MFPT versus the correlation time, τ, for the same systems as described in Figure 8 but with the variance, σ^2, set to 0.6.

the range of possible barrier heights is increased, as seen in Figure 9. In the case of the $2\mathrm{D_C}$ stochastic potential, when τ is large, all four exit channels are described by the same barrier height and hence a particle with insufficient energy to escape over one barrier can not escape over any of the others. On the other hand, in the $2\mathrm{D_I}$ stochastic potential, despite the presence of such an exit channel, the other exit channel will have a nonneglible probability of being described by a lower barrier height that gives rise to escape on a time scale shorter than τ. Consequently the incoherent potential can give rise to faster transport than the coherent potential for sufficiently large τ. As the friction is reduced and hence the particle becomes more energetically limited, this phenomenon becomes more pronounced.

To explore the effects of coherent and incoherent barriers further, as well as the impact of lattice spacing, simulations were performed for the one- and two-dimensional sinusoidal surfaces as well as the fully incoherent MHO potential. The MFPTs for these four potentials are displayed in Figure 10. Again the top, bottom, and middle panels correspond to values of the thermal friction of 0.08, 0.2, and 0.4, respectively. The MFPTs obtained from the

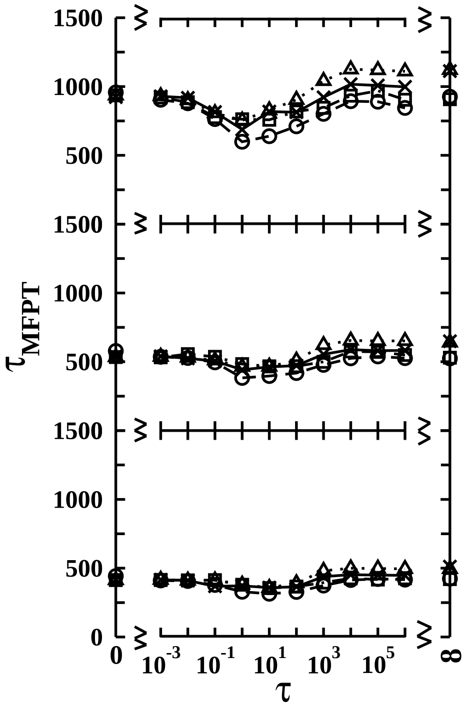

Figure 10: The MFPT for each of four stochastic potentials characterized by the variance, $\sigma^2 = 0.22$, and well-to-well distance equal to 4, but with all other parameters the same as in Figure 8. The corresponding potentials are: $1D_{\rm sin}$ (solid curve and x symbols), $2D_{\rm C}$ (dotted curve and triangles), $2D_{\rm I}$ (dot-dashed curve and squares), and $1D_{\rm MHO}$ (dashed curve and circles).

one-dimensional potentials are scaled by 1/2 as in Figure 8 and 9. The lattice spacing for all potentials is chosen as 4 and the variance is 0.22 to satisfy restrictions made by the MHO potential (see Appendix A.0.2). As opposed to Figure 8 and 9, the small η-correlation time, τ, behavior of all four potentials is equivalent. The results for the two-dimensional potentials are essentially unchanged from those of Figure 8, while the one-dimensional potentials are shifted to longer MFPTs. The only difference between Figure 8 and Figure 10 with respect to the dynamics across the corresponding stochastic potentials is an increase in the lattice spacing from $\sqrt{2}$ to 4. The resulting effects on the MFPT may be due to coupling of the two degrees of freedom in the two-dimensional potentials. Any energy acquired can be dissipated through either degree of freedom, while in one-dimension this is not possible. The two-dimensional potentials are then mostly energy diffusion limited in which the lattice spacing will have a negligible effect. However, the diffusing particles on the one-dimensional surfaces do not have access to this mechanism, and therefore spatial diffusion effects can play a small, but noticeable effect. In the large τ regime, the MFPTs increase in the order: $1\mathrm{D}_{\mathrm{MHO}} < 2\mathrm{D}_{\mathrm{I}} < 1\mathrm{D}_{\sin} < 2\mathrm{D}_{\mathrm{C}}$. In this regime, the two incoherent potentials diffuse at a faster rate than the coherent potentials as a result of the decoupled exit channels. Again, the effects of the four impassable areas of the two-dimensional potentials are reflected in their respective shifts to larger passage times.

The stochastic rate enhancement for the $1\mathrm{D}_{\sin}$, $2\mathrm{D}_{\mathrm{C}}$, and $2\mathrm{D}_{\mathrm{I}}$ potentials can be seen in Figures 11, 12, and 13, respectively, for a variance of 0.6. The top, middle, and bottom panels display the total rate, Γ_∞, single hop rate, Γ_1, and multiple hop rate, $\Gamma_m \equiv (1 - \Gamma_1)$. Each is normalized with respect to the corresponding rate obtained from the average potential $i.e.$, the $\tau = 0$ limit. The rate enhancement is seen for all three potentials, although slightly larger in one-dimension than in two. The effects of coherency play a small role as seen from the results of the $2\mathrm{D}_{\mathrm{C}}$ and $2\mathrm{D}_{\mathrm{I}}$ potentials, the latter providing a minor adjustment to the overall rate for most values of τ. However, at large τ, the rate on the coherent potential is significantly lower than the average, while the incoherent potential is able to avoid this decrease due to the uncoupled exit channels. As can be seen from all three figures, the rate enhancements for single and multiple hops are roughly equal. Although not

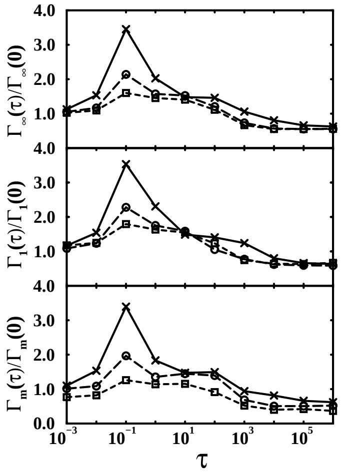

Figure 11: Hopping rates obtained on the $1D_{\rm sin}$ potential normalized with respect to the corresponding rates obtained on the average potential displayed as a function of the correlation time, τ, for a variance of $\sigma^2 = 0.6$ and values of the friction constants $\gamma_{\rm th}$: 0.08 (solid curve and x symbols), 0.2 (dashed curve and circles), and 0.4 (dotted curve and squares). The top, middle, and bottom panels display the total rate, Γ_∞, the single hopping rate, Γ_1, and the multiple hop rate, $\Gamma_{\rm m}$, respectively.

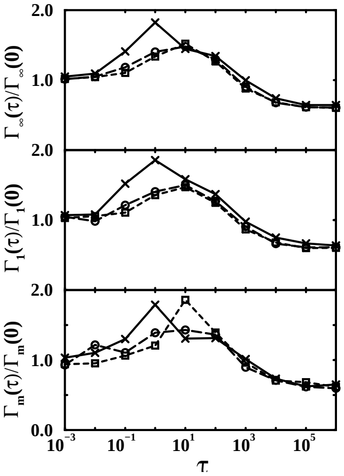

Figure 12: Normalized rates obtained from the $2D_{\rm C}$ potential displayed as a function of the correlation time, τ. All of the parameters are the same as those in Figure 11.

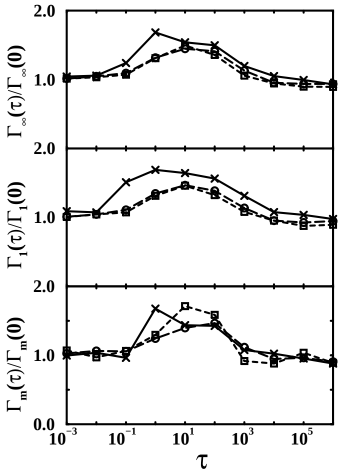

Figure 13: Normalized rates obtained from the $2\mathrm{D_I}$ potential displayed as a function of the correlation time, τ. All of the parameters are the same as those in Figure 11.

shown for brevity, the multiple hopping probabilities have also been calculated for all three potentials and follow the same trend. This is the expected result since the fluctuations in the barrier height are relatively small compared with the average and therefore should not enhance or impede a particle with sufficient energy to complete a multiple hop trajectory.

3.4 Concluding Remarks

This chapter presents an application of a recent model that describes the diffusive behavior of a Brownian particle influenced by a stochastic potential of mean force in the low friction regime. An extension into this regime was necessary to adequately characterize the complicated process of surface diffusion. Here, we have provided some insights to the fundamental nature of this process. In particular, these simulations have revealed that the dynamics on one-dimensional potentials and the corresponding two-dimensional analogue are similar, except in the extreme limits of large τ. Additionally, the results presented here are encouraging with regards to the ultimate goal of providing a new mechanism for patterning a surface by utilizing the dynamic range in the rates as modulated by τ in the barrier. This study has confirmed that the rate enhancement is observed in extended dimensions, follows a similar trend to the one-dimensional behavior, and in particular depends on τ. A forthcoming application of the model will provide a more detailed study of incoherent, but prescribed, surfaces to see the extent to which patterns can emerge on a surface [137]. The next chapter continues the discussion at the beginning of Section 3.2 on dissipating the Langevin equation in the presence of a secondary external force. In Chapter 4, an alternative method based on a space-dependent friction is constructed and a comparison between it and the method used here is presented.

CHAPTER IV

DISSIPATING THE LANGEVIN EQUATION IN THE PRESENCE OF AN EXTERNAL STOCHASTIC POTENTIAL

4.1 Introduction

In the theory of diffusion processes over fixed barriers, numerous studies have shown that the dissipative term in the Langevin equation is rarely constant along the reaction coordinate [141, 142, 143, 144, 145, 146, 147, 148]. A general rate theory when the friction is both space- and time-dependent has been developed to account for this phenomenon over the entire friction regime [149, 150, 151]. One might naïvely expect that a space-dependent component must be included in the friction kernel to capture the essential dynamics of a given system. However, this is not always the case. Several groups have shown that the average dynamical properties may still be adequately described by a generalized Langevin equation with space-independent friction even when the reaction coordinate has a strong spatial dependence [146, 142, 152, 144]. An analysis by Haynes and Voth concluded that the key factor is not whether the friction is space-dependent, since it generally will be, but rather how the friction varies along the reaction coordinate [153]. In particular, they suggest that the symmetry of the space-dependent friction with respect to the barrier can be used as a metric for evaluating the role of the friction in the dynamics. Similar product and reactant states will give rise to similar (symmetric) friction components about the transition state. Perhaps surprisingly, an antisymmetric friction does not have a significant impact on the dynamics, while a symmetric friction can result in large deviations from the predictions of standard rate theories for processes with space-independent friction [153, 146, 143, 154, 144, 148]. Thus, the Langevin model with a uniform effective friction can often approximate the dynamics of projected variables even if the formal projection would have required a space-dependent model.

The central question explored in this chapter is whether a single uniform effective friction suffices even when the Langevin system is subjected to an external space-dependent stochastic potential. As discussed extensively in the preceding chapter, the behavior of a Brownian particle diffusing across various subsets of this class of potentials has been the subject of intense research [104, 108, 39, 89, 131, 155, 156, 157, 158]. This activity has largely been motivated by the discovery of resonant activation in which the rate of transport over a stochastic barrier exhibits a maximum as a function of the correlation time in the fluctuations of the barrier height [104]. However, until recently, simulations of these systems have not been performed in the low friction regime, where deviations from equipartition may occur, due to an inability to adequately describe the friction in the presence of an additional stochastic force [109, 111]. In the preceding chapter and other previous works, the dissipation of this excess energy was achieved through a self-consistent approach in which the friction constant is renormalized iteratively until equipartition is satisfied [109, 111, 90]. This renormalization is approximate because it does not explicitly account for the correlations between the external stochastic forces across space and time, but rather uses a single mean friction to dissipate theses forces at times longer than their correlation times. A possible improvement to the self-consistent approach can be obtained by allowing the friction to be space-dependent while explicitly ignoring the memory in the stochastic potential. In the special case that the stochastic potential has no memory, then this treatment is exact. However, this approximation is often not justified when modeling real systems and therefore, the model potentials employed here are chosen to have an exponentially decaying memory of their past states. In the most extreme cases, these correlations can result in deviations from equipartition during the course of the simulation, although the space-dependent friction dissipates such fluctuations correctly in most situations. The general conclusion appears to be that the more detailed space-dependent approach is in qualitative agreement with the self-consistent approach and hence, as in the fixed barrier case, Langevin systems with stochastic forces may be dissipated by a single (though renormalized) uniform friction.

The conclusions of this chapter are supported by a study of two different classes of one-dimensional problems in which the particle diffuses across a periodic array of coherent or incoherent barriers. These two cases can be specified by the respective sinusoidal or merged-harmonic-oscillator potentials used in the preceding chapter. For such simple forms of the stochastic potential, analytic expressions for the friction as a function of the spatial coordinate can readily be obtained and are presented in Section 4.2. The resulting Langevin dynamics across these potentials dissipated either uniformly or through the space-dependent friction are illustrated in Section 4.3.

4.2 Langevin Model with Stochastic Potentials

The diffusion of a particle influenced by a stochastic potential of mean force can be adequately described by a phenomenological Langevin equation of the form,

$$\dot{v} = -\gamma v + \xi(t) + F(x;t) \ , \tag{20}$$

where $F(x;t) \equiv -\nabla_x U(x;t)$ is an external stochastic force, and γ is the friction required to dissipate both the thermal forces and those due to the external stochastic potential. As in the previous chapter the thermal bath is described by $\xi(t)$, which is a Gaussian white noise source with time correlation given by the fluctuation-dissipation relation (FDR),

$$\langle \xi(t)\xi(t') \rangle = 2k_{\mathrm{B}}T\gamma_{\mathrm{th}}\delta(t-t') \ . \tag{21}$$

In the limit that $F(x;t) = F(x;0)$ for all t, these equations reduce to the Langevin equation with $\gamma = \gamma_{\mathrm{th}}$. Otherwise, the question remains as to what is the appropriate form for the damping term. Two approaches for addressing this question are presented in Sections 4.2.2 and 4.2.3, after first describing the details of the stochastic potentials.

4.2.1 Stochastic Potential Representation

The space-dependent friction (SDF) that arises from the fluctuations in $F(x;t)$ can readily be evaluated analytically for the two classes of one-dimensional stochastic potentials outlined in the preceding chapter. The first is the sinusoidal potential described earlier by equation (15) in which the barriers fluctuate *coherently* with each other. The second is constructed using a series of merged harmonic oscillators (MHOs) given explicitly by equation (13) in which each barrier is allowed to fluctuate independently (*incoherently*) of the others. The parameters for each of the potentials are chosen such that the lattice spacing is 4 and the thermal energy of the particle is 1/6 of the average value of the barrier heights.

The stochastic term, $\eta(t)$, accounting for the time-dependent fluctuations in the barrier heights is again chosen as the Ornstein-Uhlenbeck process of the preceding chapter governed by the stochastic differential equation (16). The corresponding probability distribution is given by equation (17) with the time correlation of equation (18). The distribution of barriers heights for the sinusoidal potential is given directly by the distribution of $\eta(t)$, but due to the nature of the expression for the barrier heights of the MHOs, the resulting distribution for this potential takes on a more complex form that is sharper and slightly skewed compared with equation (17). As a result, a much smaller range of fluctuations is allowed for the MHO than the sinusoidal potential to ensure that the distribution does not become significantly non-Gaussian. Additional details on the exact behavior of the MHO barrier heights are provided in Appendix A.0.2 and reference [109]. For consistency, the dynamics of the system are again characterized by the mean first passage time approach described in Section 3.2 of the previous chapter employing a geometrical constraint in the phase space of the particle [138, 90].

4.2.2 Uniform Dissipation

In previous work [109], a self-consistent procedure was developed to ensure that the evolution of the system governed by the Langevin equation (20) remains in thermal equilibrium. This was accomplished through an iterative procedure in which the friction that is now given by the sum of the two independent contributions from the thermal bath and the stochastic potential, i.e. $\gamma \equiv \gamma_{\mathrm{th}} + \gamma_{\mathrm{F}}$, is renormalized according to the relation given in equation (12). The friction for the next iteration is determined from the value of the friction at the current iteration scaled by the magnitude of the deviation from equipartition seen in the dynamics until convergence is reached to within a desired accuracy. As mentioned in the preceding chapter, the main criticism to this approach lies in the approximation made in developing equation (12) in which the stochastic potential is treated as a local noise source, γ_{F}, obeying a fluctuation-dissipation relation equivalent to equation (21). However, the stochastic potentials have memory and are therefore nonlocal in nature leading to non-vanishing cumulants at third and higher orders. These effects are included, but only in an average manner, to second order in this approach.

4.2.3 Space-Dependent Dissipation

An alternative approach to dissipating the external stochastic force relies on replacing the space- and time-dependent friction, $\gamma(x,t)$, by a space-dependent friction, $\gamma(x(t))$, satisfying a local FDR. Given that the size of the fluctuations in $F(x;t)$ depend on x at a given t, a Brownian particle moving quickly across the surface will experience a series of forces whose relative magnitudes depend on the particle's velocity. However when the the Brownian particle moves slowly, the particle will sample only the local fluctuations of the stochastic potential in the vicinity of its local position x. In this regime, the particle arrives at a local quasi-equilibrium which must necessarily satisfy the FDR locally. This suggests that the dissipation should not be uniform, but rather should depend on position, and therefore indirectly on time. It should be noted that while the mean-field approach described in the previous subsection is capable of including the average of the correlations between the fluctuations, the approximation made here does not account for any of the memory effects. However, in the limit that there is no memory in the external stochastic potential, the following results are exact.

The question now arises of how to explicitly describe the friction constant in the presence of an additional fluctuating force resulting from the potentials of mean force given in equations (15) and (13). The friction constant must dissipate the excess energy that arises from the fluctuating forces through a local space-dependent FDR,

$$2k_{\mathrm{B}}T\gamma_{\mathrm{c}}(x;t) = \langle \delta F_{\mathrm{c}}(x;t)^2 \rangle \ , \tag{22}$$

where the cumulative force is simply the sum of the thermal Gaussian noise and the stochastic force arising from the external potential, $F_{\mathrm{c}} = F_{\mathrm{th}} + F_{\mathrm{U}}$. Assuming the respective fluctuations in the bath and the potential are uncorrelated, *i.e.* $\langle \delta F_{\mathrm{th}} \delta F_{\mathrm{U}} \rangle = 0$, then equation (22), reduces to

$$2k_{\mathrm{B}}T\gamma_{\mathrm{c}}(x;t) = \langle \xi(t)\xi(t') \rangle + \langle \delta F_{\mathrm{U}}(x;t)^2 \rangle \ , \tag{23}$$

The thermal fluctuations are Ohmic as given in equation (21), and the relationship for the fluctuations in the force is $\delta F_{\mathrm{U}}(x;t) \equiv F_{\mathrm{U}}(x;t) - \langle F_{\mathrm{U}}(x;t)\rangle_{\eta}$, where the average is taken with respect to the auxiliary stochastic variable, η. The average value of the force can be determined according to the usual integrals,

$$\langle F_{\mathrm{U}}(x;t)\rangle = \frac{-\int_{-\infty}^{\infty} d\eta\, P(\eta)\nabla_x U(x;t)}{\int_{-\infty}^{\infty} d\eta\, P(\eta)} \; , \tag{24}$$

where the fluctuations in the force are governed by the stochastic Ornstein-Uhlenbeck process, η, whose probability distribution is given by equation (17).

The remaining steps of the derivation rely upon the specific form of the potential. As an illustration, the SDF is evaluated explicitly below for the simpler sinusoidal (coherent) stochastic potential. (The results for the incoherent MHO potential can be found in Appendix B.) The derivation begins by direct evaluation of equation (23) for the specific class of potentials. As remarked above, the first term reproduces the FDR, equation 21, for the thermal forces. Ignoring the correlation in the forces at different times, the second reduces to:

$$\begin{aligned}\langle \delta F_{\mathrm{U}}(x;t)^2\rangle \quad &= \quad \frac{\pi^2}{4}\cos^2\left(\frac{\pi x}{2}\right)\int_{-\infty}^{\infty} d\eta \left(E_b + \frac{1}{2}\eta\right)^2 P(\eta) - &(25)\\ &\quad\;\; \left[\frac{\pi}{2}\cos\left(\frac{\pi x}{2}\right)\int_{-\infty}^{\infty} d\eta \left(E_b + \frac{1}{2}\eta\right) P(\eta)\right]^2 \; . &(26)\end{aligned}$$

The Gaussian integrals are readily evaluated to yield:

$$\langle \delta F_{\mathrm{U}}(x;t)^2\rangle = \frac{\sigma^2\pi^2}{16}\cos^2\left(\frac{\pi x}{2}\right) \; . \tag{27}$$

Upon substitution into equation (22), the explicit form of the SDF is

$$\gamma_{\mathrm{c}}(x;t) = \gamma_{\mathrm{th}}\delta(t-t') + \frac{\sigma^2\pi^2}{32 k_{\mathrm{B}}T}\cos^2\left(\frac{\pi x}{2}\right) \; . \tag{28}$$

This is the simplest possible form for this result, and is due to the separability of the potential into a sum of deterministic and linear stochastic terms. In fact, it is easily shown that for any separable potential of the form,

$$U(x;t) = \bar{U}(x) + \eta(t) W(x) \; , \tag{29}$$

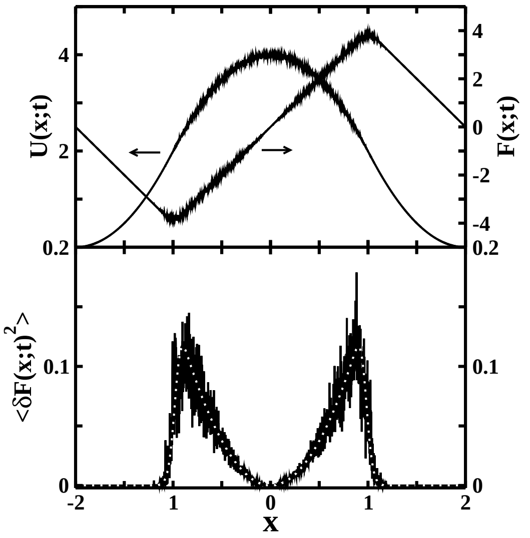

Figure 14: Representative fluctuations over one period of the MHO potential and force (top panel), and the resulting space-dependent friction (bottom panel). The numerical component in the bottom panel is displayed as the solid black line, with the analytic result, given in Appendix B, as the dotted white line. The temperature is 2/3, the variance is 0.22, and the thermal friction is 0.08.

where $\bar{U}(x)$ is the deterministic component of the potential of mean force, then the additional friction due to the stochastic potential is given by

$$\langle \delta F_{\mathrm{U}}(x;t)^2 \rangle = (\nabla_x W(x))^2 \int d\eta \, (\eta^2 - \eta) P(\eta) \ , \tag{30}$$

provided the distribution is normalized. The MHO does not satisfy the condition of equation (29) and hence its friction correction can not be obtained by equation (30). The form of the friction correction for the MHO consequently contains more terms, but the requisite approximation (that the forces are uncorrelated at different times) enters the derivation in a conceptually equivalent way.

4.3 Results and Discussion

The analytic and numerical space-dependent components of the friction over one period of the MHO and sinusoidal potentials can be seen in the bottom panels of Figures 14 and 15, respectively, with the numerical results averaged over 500 representative trajectories. The top panels display the fluctuations in the potential and the resulting forces that give rise to the space-dependent friction. The analytic forms of the SDF, displayed as the dotted white

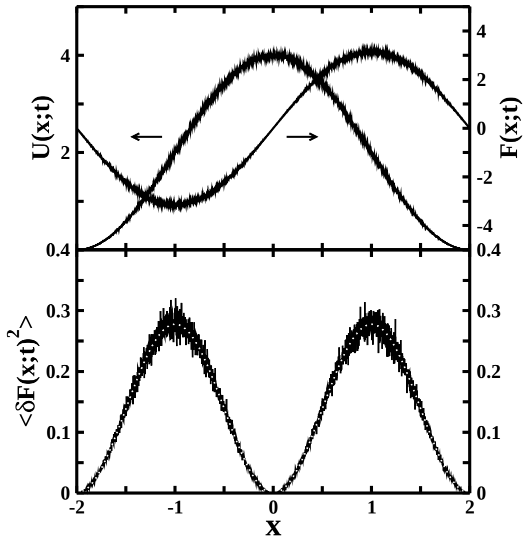

Figure 15: Representative fluctuations over one period of the sinusoidal potential and force (top panel), and the resulting space-dependent friction (bottom panel). The numerical result is displayed as the solid black line, with the analytic result, given by equation (28), shown as the dotted white line. The parameters used are the same as in Figure 14.

line, agree with the corresponding numerical results, and exact agreement is obtained upon further averaging. The fluctuations in the forces reach a maximum at approximately the midpoint between the minima and maxima, where deviations from the average force take on the largest values. The fluctuations in the potential are largest at the barriers, while the forces are zero at these locations. This leads to a vanishing contribution to the total friction from the space-dependent component at these points. In the well region, the behavior of the SDF for the sinusoidal and MHO potentials is inherently different. The SDF for the MHO is zero outside of the barrier region since the wells do not fluctuate by construction. However, the sinusoidal potential fluctuates continuously throughout leading to a friction correction along the entire reaction coordinate. Consequently, the magnitude of the friction correction in simulations employing the sinusoidal potential are slightly larger than that in those employing the MHO. But, as illustrated below, this effect does not have a dramatic effect on the resulting dynamics.

Values of the friction corrections calculated from the iterative and space-dependent approaches for the MHO and sinusoidal potentials are displayed in Table 1 with the values of the thermal friction listed in the left-most column. The variance and correlation time for

Table 1: The average of the friction corrections, γ_{F}, calculated by the iterative self-consistent (0) and space-dependent (sdf) approaches for the MHO and sinusoidal potentials. The resulting temperatures are also included for the space-dependent friction. In all cases the temperature is 2/3 (in units of a standard temperature, $k_{\mathrm{B}}T_0$), the variance, $\sigma^2 = 0.22$, and the correlation time, $\tau_{\mathrm{c}} = 1$.

	MHO			Sin		
γ_{th}	$\langle\gamma_{\mathrm{F}}\rangle_0$	$\langle\gamma_{\mathrm{F}}\rangle_{\mathrm{sdf}}$	$\langle v^2\rangle_{\mathrm{sdf}}$	$\langle\gamma_{\mathrm{F}}\rangle_0$	$\langle\gamma_{\mathrm{F}}\rangle_{\mathrm{sdf}}$	$\langle v^2\rangle_{\mathrm{sdf}}$
0.08	0.00	0.01	0.67	0.00	0.03	0.69
0.2	0.00	0.01	0.67	0.00	0.03	0.68
0.4	0.00	0.01	0.67	0.01	0.03	0.67

both potentials is 0.22 and 1, respectively. The resulting temperatures, ($k_{\mathrm{B}}T \equiv \langle v^2\rangle$), are also listed for the space-dependent approach. The friction correction in the self-consistent method ensures equipartition by definition, and therefore, is not listed. The magnitude of the SDF for all values of τ_{c} follow accordingly; however this is the only value with respect to the given variance for which any deviation from equipartition is observed. As can be seen, both the self-consistent and space-dependent components of the total friction for each potential provide negligible contributions for this variance since the magnitude of the fluctuations in the barrier height are relatively small. Therefore the total friction is a sum of a large thermal component and a space-dependent contribution. The slight differences in the magnitudes of the SDF for the two potentials can be attributed to the piecewise nature of the MHO potential. The particles spend most of the simulation time in the wells which do not fluctuate. A contribution to the total friction from the space-dependent term is included only when the energetically-limited particle accumulates enough energy to explore the upper portion of the MHO potential.

To further explore the accuracy of the space-dependent approach, the sinusoidal potential has been studied with a ten-fold increase in the variance from 0.22 to 2.2. The values of the friction correction from these simulations are listed in Table 2.

Table 2: The average of the friction corrections, γ_F, calculated by the iterative self-consistent (0) and space-dependent approaches (sdf) for the sinusoidal potential. The resulting temperatures are also included for the space-dependent friction method. The temperature is 2/3 in all cases and the variance, $\sigma^2 = 2.2$.

	$\tau_\mathrm{c} = 10^{-1}$			$\tau_\mathrm{c} = 10^{0}$			$\tau_\mathrm{c} = 10^{1}$		
γ_th	$\langle\gamma_\mathrm{F}\rangle_0$	$\langle\gamma_\mathrm{F}\rangle_\mathrm{sdf}$	$\langle v^2\rangle_\mathrm{sdf}$	$\langle\gamma_\mathrm{F}\rangle_0$	$\langle\gamma_\mathrm{F}\rangle_\mathrm{sdf}$	$\langle v^2\rangle_\mathrm{sdf}$	$\langle\gamma_\mathrm{F}\rangle_0$	$\langle\gamma_\mathrm{F}\rangle_\mathrm{sdf}$	$\langle v^2\rangle_\mathrm{sdf}$
0.08	0.04	0.28	0.72	0.05	0.29	0.74	0.01	0.28	0.68
0.2	0.04	0.28	0.71	0.05	0.29	0.72	0.01	0.28	0.68
0.4	0.04	0.28	0.70	0.06	0.29	0.71	0.01	0.28	0.67

The displayed correlation times, τ_c, are those that exhibit the largest resonant activation. Consequently, if memory effects in the barrier heights are important in determining the friction constant, it should be manifested here. Although not shown for brevity, outside this region of the correlation time, the magnitude of the deviations from equipartition decrease rapidly, but the size of the space-dependent components remains roughly constant. Similarly, the corresponding corrections arising in the self-consistent method also approach zero. As can be seen from Table 2, the space-dependent approach results in a correction that is roughly constant for all values of the correlation time, while the iterative approach does exhibit some variation with τ_c. This is the expected result since the space-dependent friction assumes the fluctuations in the potential are local and therefore, ignores any correlation in the barrier heights. The iterative approach, however, is capable of incorporating the memory of the potential into the friction correction, but only in an average manner. As a consequence, significant deviations from equipartition may be observed when simulations are performed with a space-dependent friction that ignores the correlation effects, as illustrated by this extreme example.

Figures 16 and 17 display the MFPTs obtained for the MHO potential with the results from the space-dependent and self-consistent approaches in the top and bottom panels, respectively. The results in Figure 16 have been calculated using a variance of $\sigma^2 = 0.05$, while those in Figure 17 use $\sigma^2 = 0.22$. The corresponding results for the sinusoidal potential using a variance of 0.22 can be seen in Figure 18. The values on the broken axis represent the numerically calculated MFPTs at the limits of the correlation time, τ_c. In the zero-correlation time limit, the fluctuations in the potential are so rapid that the particle effectively experiences the average, stationary potential, from which the dynamics were calculated. In the limit of infinite correlations, fluctuations in the potential are nonexistent, and therefore the particle experiences a single realization of the potential with constant barrier heights determined by the initial value sampled from the distribution. The MFPTs displayed in Figure 17 obtained with a larger variance alters the magnitude of the resonant activation, but influences the results for the two approaches equally. The results from the simulations with a space-dependent friction are systematically shifted to lower MFPTs as

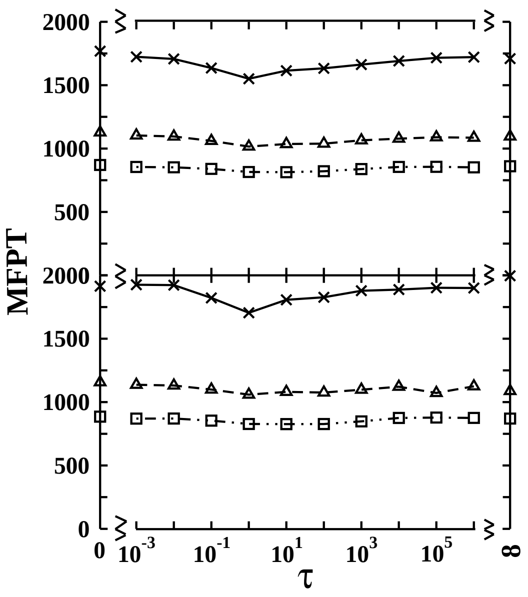

Figure 16: The mean first-passage times (MFPT) for a particle diffusing across the MHO stochastic potential are displayed for two possible scenarios of the dissipative mechanism. The top panel uses space-dependent friction, and the bottom displays the uniform friction determined by the self-consistent method. The variance for both is 0.05, and the three lines correspond to values of the thermal friction of 0.08 (solid curve with x symbols), 0.2 (dashed curve with triangles), and 0.4 (dot-dashed curve with squares). The symbols on the broken axis represent the numerically calculated MFPTs at the limits of the correlation time.

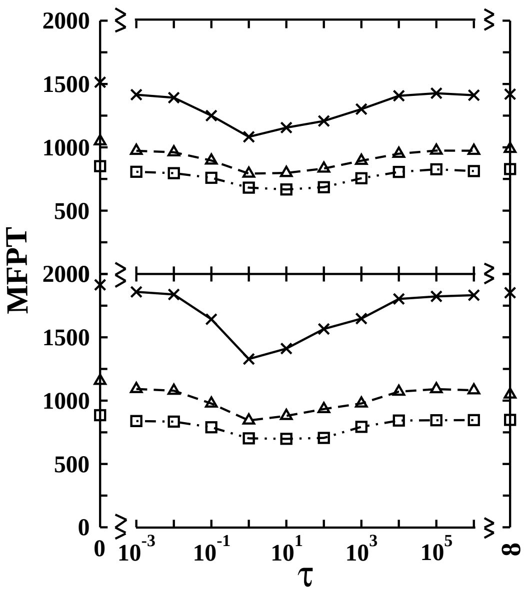

Figure 17: The mean first-passage times (MFPT) for a particle diffusing across the MHO stochastic potential are displayed for two possible scenarios of the dissipative mechanism. The parameters are the same as in Figure 16, except the variance is 0.22.

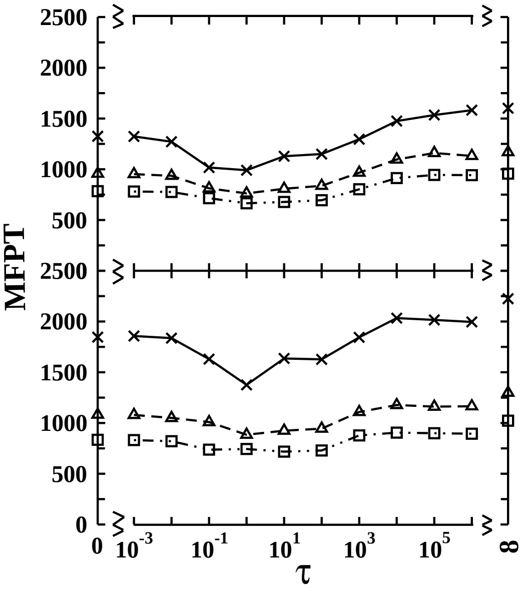

Figure 18: The mean first-passage times (MFPT) for a particle diffusing across the sinusoidal stochastic potential are displayed for two possible scenarios of the dissipative mechanism. Other than the change from the MHO to the sinusoidal potential, the parameters are the same as in Figure 17.

seen in all three figures. This trend is most readily explained through Table. 1. In the low friction regime, an increase in the friction increases the corresponding rate of transport. The average space-dependent contribution is always larger than its respective mean field counterpart, and is expected to have the largest effect on the results with the smallest thermal friction. The fluctuations present along the entire reaction coordinate of the sinusoidal potential do not appear to have a dramatic effect on the dynamics. The results in Figure 18 for the sinusoidal potential follow the same trend as those in Figure 16 and 17 for the MHO potential indicating that the SDF approach is capable of adequately describing the fluctuations in the system. Aside from the shift, the general behavior of the MFPT is adequately reproduced by both methods, particularly at larger values of the thermal friction when the space-dependent component becomes less significant. At this level of description, each of the two approaches for constructing the friction are capable of capturing the essential dynamics of the system. However, some advantage is gained by using the self-consistent method because it ensures the system is kept at constant temperature for all values of the correlation time throughout the simulation, while the space-dependent approach may lead to deviations in extreme cases. The most significant difference between the two methods can be seen at intermediate correlation times, in which the resonant activation observed from the iterative approach is slightly more pronounced. This can particularly be seen in the MFPTs when the friction takes on the smallest value of $\gamma_{\mathrm{th}} = 0.08$. Since the resonant activation arises from correlations in the barrier heights, it is not surprising that simulations incorporating a friction capable of accounting for this phenomenon can have a noticeable impact on the dynamics, even if it does so only in an average manner.

4.4 ***Conclusions***

The space-dependent friction arising from the presence of a secondary (external) stochastic potential in the Langevin equation has been explicitly derived for two simple classes of stochastic potentials. The numerical results are in excellent agreement with analytic expressions describing the space-dependent friction. The resulting dynamics have been compared to those obtained using an alternate approach in which a uniform correction is calculated self-consistently. Although the latter approach does effectively include the time correlation between the barrier fluctuations at long times, the former does not in any sense. This neglect may result in deviations from equipartition in some extreme cases. However, both approaches are capable of capturing the essential dynamics of the system and lead to the now-expected resonant activation phenomenon. Consequently, the central result of this chapter is that the Langevin dynamics of a particle under the influence of external stochastic potentials can be properly dissipated by a single uniform renormalized friction without loss of qualitative (and often quantitative) accuracy.

The role of the memory time in an external stochastic potential acting on a particle described by a generalized Langevin equation of motion is still an open question. In this limit, there would presumably be an interplay between the memory time of the thermal friction and that of the stochastic potential. When the latter is small compared to the former, the quasi-equilibrium condition central to this work would no longer be satisfied by the particle, and hence it is expected that a non-uniform (and time-dependent) friction correction would then be needed.

This concludes the explicit work on describing surface diffusion through molecular dynamics and Langevin simulations, although the following chapter is implicitly connected to all the results described thus far. In Chapter 5 an alternative method to calculate chemical reaction rates based on a moving dividing surface is discussed. This quantity has been the primary observable in the studies of diffusion on metal surfaces and over fluctuating barriers described in the previous chapters although rate estimates are generally of importance to practically all molecular systems.

CHAPTER V

IDENTIFYING REACTIVE TRAJECTORIES USING A MOVING TRANSITION STATE

5.1 Introduction

It is perhaps surprising how many problems in chemistry, physics, and biology can be reduced to the simple model of diffusion over a barrier [159]. Although chemical reactions in all phases of matter provide the prime example [160, 161], a plethora of systems that evolve from suitably defined "reactant" to "product" states are amenable to a description in this framework [162, 163, 164, 165, 166, 167]. The diffusion of adsorbates across a surface discussed in Chapters 2, 3 and 4 necessarily satisfies this criteria. Transition State Theory (TST) [168, 47, 161] or one of its descendants [169, 125, 124] is often used to approximate the rate of these reaction processes. This theory is based on the assumption that the reaction rate is determined in a small volume of the phase space near the barrier. It is then possible to define a dividing surface separating reactants from products and obtain the rate from the flux through this surface. The optimal location of the dividing surface is that which minimizes the number of recrossings —the fundamental idea of variational transition state theory [170, 171, 172, 173, 174]. When the system of interest can be viewed as isolated from its environment, as in low-density gas phase chemical reactions, TST may indeed provide an excellent approximation to the rate. However, most processes of interest do not take place in isolation, but rather in a complex environment where interactions between the system and its surroundings occur on time scales comparable to that of the reaction: In a reaction occurring in the condensed phase, for instance, the dynamics of the solute is strongly coupled to that of the solvent. In this case, the fundamental assumption of TST, that the dividing surface is crossed once and only once, no longer holds [175, 176, 142, 177]. On the time scale of the reaction event, fluctuations of the environment will almost inevitably cause recrossings of the dividing surface that lead to an overestimation of the rate.

To overcome the recrossing problem, many TS dividing surfaces have been suggested in the literature [178, 161, 179, 174] which provide systematic (and simple) approximations to the optimal TS dividing surface. In some cases, the dividing surface has been identified in the infinite-dimensional phase space consisting of the system and an explicit set of bath oscillators [180, 181, 179]. This approach leads to an excellent approximation to the rate [179, 125]. Interestingly, the same result was subsequently obtained without recourse to the explicit heat bath model, using instead a collective reaction coordinate containing the influence of the bath directly [182].

In a recent series of papers [183, 184], the recrossing problem was reformulated using a dividing surface that is itself moving stochastically so as to avoid recrossings. The motion of that surface follows the unique trajectory —named the TS trajectory— that never leaves the barrier region. Any reactive trajectory crosses the moving surface once and only once, whereas a nonreactive trajectory does not cross at all. This construction extends the approach of [182] in that it provides not only a reaction coordinate, but also the complete geometric structure by specifying all of the unstable and stable degrees of freedom globally [185]. The previous purely analytic studies [183, 184] are complemented here with a numerical investigation of the reaction dynamics for a two-dimensional stochastic nonlinear model. It will be shown that the moving dividing surface offers considerable computational advantages over the traditional fixed surface: Its use can significantly reduce the simulation time required to distinguish between reactive and nonreactive trajectories. Indeed, for a harmonic barrier it identifies reactive trajectories *a priori*, so that the need to simulate their dynamics does not arise at all. In an anharmonic system, the identification of reactive trajectories by the moving surface is no longer exact. Nevertheless, for moderately strong anharmonicities it provides a useful approximation, and its advantages over the fixed surface are retained. In addition, the moving TS surface introduces novel observables that characterize the reaction process on a microscopic level. Most prominently, it allows one to define a unique reaction time for each individual trajectory.

The outline of this chapter is as follows: In Section 5.2, the two-dimensional stochastic nonlinear model that is the focus of the computational studies in this chapter is defined and the construction of the moving TS dividing surface and its associated geometric structures is briefly reviewed. In Section 5.3, the ensemble of trajectories is specified by a thermal distribution of particles localized at the conventional TST dividing surface. This barrier ensemble is reminiscent of the weighting distribution in standard rate expressions and is appropriate even in nonlinear cases. Its simple structure also readily leads to the analytic determination of several observables of the model system in the harmonic limit (Section 5.4). They are in precise agreement with the numerical results presented in Section 5.5. The latter section also demonstrates that observables converge faster when evaluated using the moving dividing surface rather than conventional numerical methods, both in the harmonic limit and in systems with anharmonic barriers. This observation is particularly useful in the anharmonic case when the chosen system is not amenable to analytic approaches.

5.2 Preliminaries

Although the general theory is applicable to systems with an arbitrary number n of degrees of freedom, the following discussion will be restricted to $n = 2$ coordinates under the influence of a stochastic bath. This choice can be made without loss of generality because it exhibits all the salient features of the higher-dimensional cases: It can encompass an unbound (reactive) direction and a bound bath mode whose interaction with the reactive mode is strong enough to require its explicit description. The coupling of the modes is described by a Taylor expansion about a transition point (or col) on the potential energy surface. Such a model with a minimum number of nonlinear terms is described in this section. It will be used in the following to study the effect of increasing anharmonicity on the identification of reactive trajectories.

To set the stage for the following investigations, the construction of the moving TST dividing surface and the associated invariant manifolds is summarized in the remainder of this section. The reader interested in a full exposition is referred to references [183] and [184], where the formalism was first introduced.

5.2.1 The Two-Dimensional Dissipative Model

A prototypical reactive system within a solvent may be described by the Langevin equation [84]

$$\ddot{\vec{q}}_\alpha(t) = -\nabla_{\vec{q}} U(\vec{q}_\alpha(t)) - \mathbf{\Gamma}\dot{\vec{q}}_\alpha(t) + \vec{\xi}_\alpha(t) \ . \tag{31}$$

The vector $\vec{q}$ here denotes a set of $n = 2$ mass-weighted coordinates, $U(\vec{q})$ the potential of mean force governing the reaction, $\mathbf{\Gamma}$ a symmetric positive-definite friction matrix and $\vec{\xi}_\alpha(t)$ a fluctuating force assumed to be Gaussian with zero mean. The subscript α represents randomness by labeling different instances of the fluctuating force. The latter is related to the friction matrix $\mathbf{\Gamma}$ by the fluctuation-dissipation theorem [84]

$$\left\langle \vec{\xi}_\alpha(t)\vec{\xi}_\alpha^{\mathrm{T}}(t') \right\rangle_\alpha = 2\, k_{\mathrm{B}} T\, \mathbf{\Gamma}\, \delta(t-t') \ , \tag{32}$$

where the angular brackets denote the average over the instances α of the noise. Although not strictly necessary, the friction is often taken to be isotropic, *i.e.*,

$$\mathbf{\Gamma} = \gamma \boldsymbol{I} \ , \tag{33}$$

with a scalar friction constant γ.

The reactant and product regions in configuration space are separated by a potential barrier whose position is marked by a saddle point $\vec{q}_0^{\ddagger} = 0$ of the potential $U(\vec{q})$. In its vicinity, the potential is approximately harmonic and can always be written in a diagonal normal form. In general, anharmonic terms will be present in the potential. In the neighborhood of the saddle point, where the reaction rate is determined, they are only moderately strong, but usually not negligible. In this chapter, we include a typical (even) higher-order nonlinear term and focus on the potential

$$U(x, y) = -\frac{1}{2}\omega_x^2 x^2 + \frac{1}{2}\omega_y^2 y^2 + k x^2 y^2 \ , \tag{34}$$

where the position vector is written as $\vec{q} = (x, y)$, and the constant k quantifies the nonlinear coupling of the different degrees of freedom. Note that the nonlinearity in the potential (34) is symmetric in the coordinate system and neglects other fourth order terms that are typically retained in the analysis of anharmonic barriers. (See, e.g., reference [186], in which such coupled anharmonic potentials have been used to study the $\mathrm{H}+\mathrm{H}_2 \rightleftharpoons \mathrm{H}_2+\mathrm{H}$ reaction and bound vibrational systems.) However, as discussed in Appendix C, it is amenable to an analytic treatment that simplifies the numerical computation of the forward and backward trajectories, while providing sufficient coupling to break the exact integrability of the harmonic system.

In the special case $k = 0$, the system is globally harmonic. In this instance, the constructions outlined below yield a moving dividing surface that is strictly free of recrossings. If $k \neq 0$, deviations from the harmonic dynamics will arise outside the TS region that may lead to error in the identification of reactive trajectories. Nonetheless, the wealth of microscopic detail that the moving dividing surface reveals can most easily be illustrated using the harmonic limit. This is shown in Sections 5.3 and 5.5.1. The real power of the numerical method, however, lies in addressing nonlinear systems; the accuracy of the approximate identification of nonlinear reactive trajectories is discussed in Section 5.5.2.

With the potential (34), the Langevin equation (31) reads

$$\ddot{\vec{q}}_\alpha(t) = \mathbf{\Omega}\vec{q}_\alpha(t) + O(q_\alpha^3) - \mathbf{\Gamma}\dot{\vec{q}}_\alpha(t) + \vec{\xi}_\alpha(t) \; , \tag{35}$$

where

$$\mathbf{\Omega} = \begin{pmatrix} \omega_x^2 & 0 \\ 0 & -\omega_y^2 \end{pmatrix} \tag{36}$$

is the matrix of second derivatives of $U(\vec{q})$. The nonlinear terms in equation (35), which stem from the anharmonic contributions to the potential (34) will be ignored in the remainder of this section, where an exact dividing surface for the harmonic limit will be constructed. The full nonlinear equation of motion (31) will be taken up again in the numerical calculations of Section 5.5.2. The following presentation can easily be generalized to N spatial dimensions if y is understood to denote an $(n-1)$-dimensional vector and the corresponding squared frequency ω_y^2 an $(n-1)$-dimensional symmetric matrix whose eigenvalues need not be degenerate.

5.2.2 The Transition State Trajectory

As was shown in [183, 184], equation (35) can be rewritten in phase space, $\vec{z} = (\vec{q}, \vec{v})$, with $\vec{v} = \dot{\vec{q}}$, as

$$\dot{\vec{z}}_\alpha(t) = A\vec{z}_\alpha(t) + \begin{pmatrix} 0 \\ \vec{\xi}_\alpha(t) \end{pmatrix} \tag{37}$$

with the $2n$-dimensional constant matrix

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{0} & \boldsymbol{I} \\ \boldsymbol{\Omega} & -\boldsymbol{\Gamma} \end{pmatrix} , \tag{38}$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix. The matrix $\boldsymbol{A}$ is readily diagonalized to yield the eigenvalues ϵ_j and the corresponding eigenvectors $\vec{V}_j$. Equation (37) then decomposes into a set of $2n$ independent scalar equations of motion

$$\dot{z}_{\alpha j}(t) = \epsilon_j z_{\alpha j}(t) + \xi_{\alpha j}(t) , \tag{39}$$

where $z_{\alpha j}$ are the components of $\vec{z}$ in the basis $\vec{V}_j$ of eigenvectors of $\boldsymbol{A}$ and $\xi_{\alpha j}$ are the corresponding components of $(0, \vec{\xi}_\alpha(t))$.

A particular solution of equation (39) is given by

$$z^{\ddagger}_{\alpha j}(t) = \begin{cases} \int_{-\infty}^{0} e^{-\epsilon_j \tau} \xi_{\alpha j}(t+\tau)\, d\tau & \text{if } j \text{ such that } \operatorname{Re}\epsilon_j < 0, \\ -\int_{0}^{\infty} e^{-\epsilon_j \tau} \xi_{\alpha j}(t+\tau)\, d\tau & \text{if } j \text{ such that } \operatorname{Re}\epsilon_j > 0. \end{cases} \tag{40}$$

Whereas a typical trajectory will eventually descend into either the reactant or the product wells, the trajectory given by equation (40) has the important property [183, 184] that it remains in the vicinity of the saddle point for all time. In this respect it resembles the equilibrium position on the saddle that represents the unique trajectory in the absence of noise that never descends from the saddle. We named this distinguished trajectory the Transition State Trajectory in [183, 184] because it plays as central a role in the TST in a noisy environment as the equilibrium point does in conventional TST. Although the integral representation (40) defines the TS trajectory, it does not provide the most efficient way of calculating it. In fact, by means of an algorithm that was introduced in [184] an instance of the TS trajectory can be sampled almost as efficiently as an instance of the fluctuating force itself.

5.2.3 The Relative Dynamics

Once the TS Trajectory $\vec{z}^{\ddagger}_{\alpha}(t) = (\vec{q}^{\ddagger}_{\alpha}(t), \vec{v}^{\ddagger}_{\alpha}(t))$ is given, any other trajectory under the influence of the same noise can be described in relative coordinates

$$\Delta\vec{z}(t) = \begin{pmatrix} \Delta\vec{q}(t) \\ \Delta\vec{v}(t) \end{pmatrix} = \vec{z}_{\alpha}(t) - \vec{z}^{\ddagger}_{\alpha}(t) \;, \tag{41}$$

where the TS Trajectory serves as the origin of a moving coordinate system. The relative coordinate vectors can be written without a subscript α because they satisfy the noiseless equation of motion

$$\Delta\ddot{\vec{q}}(t) = \mathbf{\Omega}\, \Delta\vec{q}(t) - \mathbf{\Gamma}\, \Delta\dot{\vec{q}}(t) \;, \tag{42}$$

or, in phase space,

$$\Delta\dot{\vec{z}}(t) = \boldsymbol{A}\, \Delta\vec{z}(t) \tag{43}$$

and are, therefore, independent of the noise. Using the eigenvectors of $\boldsymbol{A}$, one can construct invariant manifolds and a no-recrossing surface of the noiseless relative dynamics. According to equation (41), they can then be regarded as being attached to the TS Trajectory and being carried around by it. In this way one obtains moving invariant manifolds and a moving no-recrossing surface in the phase space of the full, noisy dynamics [183, 184].

In the two-dimensional case of the potential in equation (34) under isotropic friction as specified in equation (33), the eigenvalues of $\boldsymbol{A}$ can be found explicitly:

$$\begin{aligned} \epsilon_{\mathrm{u}} &= -\frac{1}{2}\left(\gamma - \sqrt{\gamma^2 + 4\omega_x^2}\right) \;, & \epsilon_{\mathrm{s}} &= -\frac{1}{2}\left(\gamma + \sqrt{\gamma^2 + 4\omega_x^2}\right) \;, \\ \epsilon_{\mathrm{t1}} &= -\frac{1}{2}\left(\gamma - \sqrt{\gamma^2 - 4\omega_y^2}\right) \;, & \epsilon_{\mathrm{t2}} &= -\frac{1}{2}\left(\gamma + \sqrt{\gamma^2 - 4\omega_y^2}\right) \;. \end{aligned} \tag{44}$$

The corresponding eigenvectors read

$$\vec{V}_{\mathrm{u}} = \begin{pmatrix} 1 \\ 0 \\ \epsilon_{\mathrm{u}} \\ 0 \end{pmatrix} \;, \qquad \vec{V}_{\mathrm{s}} = \begin{pmatrix} 1 \\ 0 \\ \epsilon_{\mathrm{s}} \\ 0 \end{pmatrix} \;, \qquad \vec{V}_{\mathrm{t1}} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \epsilon_{\mathrm{t1}} \end{pmatrix} \;, \qquad \vec{V}_{\mathrm{t2}} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \epsilon_{\mathrm{t2}} \end{pmatrix} \;. \tag{45}$$

These simple analytic results are obtained because isotropic friction leads to a decoupling of the reactive and the transverse degrees of freedom. The eigenvectors $\vec{V}_\mathrm{u}$ and $\vec{V}_\mathrm{s}$ span the reactive x-v_x subspace, whereas $\vec{V}_{\mathrm{t}1}$ and $\vec{V}_{\mathrm{t}2}$ span the transverse subspace y-v_y.

The knowledge of the eigenvectors allows one to explicitly specify the coordinate transformation between position-velocity coordinates $\Delta x, \Delta y, \Delta v_x, \Delta v_y$ and the diagonal coordinates $\Delta x_\mathrm{u}, \Delta x_\mathrm{s}, \Delta x_{\mathrm{t}1}, \Delta x_{\mathrm{t}2}$ that characterize a phase space point via $\Delta\vec{x} = \sum_i \Delta x_i \vec{V}_i$. In the reactive subspace, these transformations read

$$\Delta x = \Delta x_\mathrm{u} + \Delta x_\mathrm{s}\ , \qquad \Delta v_x = \epsilon_\mathrm{u}\,\Delta x_\mathrm{u} + \epsilon_\mathrm{s}\,\Delta x_\mathrm{s} \tag{46}$$

with the inverse

$$\Delta x_\mathrm{u} = \frac{\Delta v_x - \epsilon_\mathrm{s}\,\Delta x}{\epsilon_\mathrm{u} - \epsilon_\mathrm{s}}\ , \qquad \Delta x_\mathrm{s} = \frac{-\Delta v_x + \epsilon_\mathrm{u}\,\Delta x}{\epsilon_\mathrm{u} - \epsilon_\mathrm{s}}\ . \tag{47}$$

For all values of γ and ω_x, the eigenvalue ϵ_u is positive, whereas ϵ_s is negative. They correspond to one-dimensional stable and unstable subspaces within the reactive degree of freedom which, together with representative trajectories, are illustrated in Figure 19. The coordinate Δx_u determines the fate of a trajectory in the remote future: Trajectories with $\Delta x_\mathrm{u} > 0$ descend into the product well, those with $\Delta x_\mathrm{u} < 0$ into the reactant well. Similarly, a stable coordinate $\Delta x_\mathrm{s} > 0$ indicates a trajectory that comes out of the product well in the distant past, whereas a trajectory with $\Delta x_\mathrm{s} < 0$ comes out of the reactant well. A forward-reactive trajectory that changes from reactants to products is thus characterized by $\Delta x_\mathrm{s} < 0$ and $\Delta x_\mathrm{u} > 0$, whereas a backward-reactive trajectory has $\Delta x_\mathrm{s} > 0$ and $\Delta x_\mathrm{u} < 0$. Each reactive trajectory crosses the line $\Delta x = 0$ once and only once. This line, or in several degrees of freedom the hypersurface defined by this condition, can therefore serve as a recrossing-free dividing surface between reactants and products. Furthermore, the invariant stable and unstable subspaces themselves act as separatrices between reactive and nonreactive trajectories. Once the initial condition of a trajectory is known relative to these separatrices, it can unambiguously be classified as reactive or non-reactive.

5.3 The Barrier Ensemble

The rate calculation of infrequent events —such as that in chemical reactions— can be greatly simplified by sampling trajectories in the transition state region rather than in

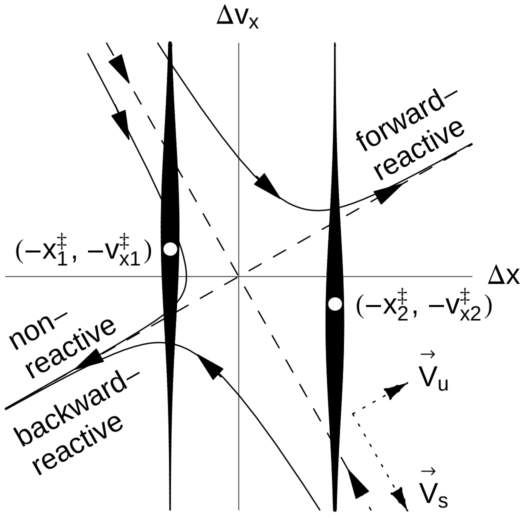

Figure 19: Phase portrait illustrating the relative dynamics in the reactive degree of freedom. Dashed lines indicate the stable and unstable manifolds of the equilibrium point, and the dotted arrows display the corresponding eigenvectors that span the diagonal coordinate system. Solid curves illustrate representative trajectories. White dots indicate two possibilities for the position and velocity of the TS trajectory at time $t = 0$, the vertical lines the corresponding barrier ensembles. The probability density is given by the line widths.

the reactant region [187, 188, 44, 189, 190]. The transition path sampling technique, for example, focuses exclusively on reactive trajectories and therefore mitigates the difficulty of studying high-dimensional systems [191, 192, 193, 194, 195]. Nonetheless, the rate of infrequent events has long been known to be described by a flux-flux correlation function relative to a fixed dividing surface [196, 197, 44]. (But see reference [198] for a recent enhanced-sampling strategy to smooth the potential and thereby speed up the calculations.) One difficulty in computing the correlation function, however, is the need for the simulated trajectories to be evolved for very long times simply to determine which trajectories are reactive. We will show below that the use of the time-dependent TST dividing surface may allow one to resolve that question in a more efficient way by identifying the nature of the trajectory —*viz.* reactive or not— at significantly earlier evolution times.

In order to sample the reactive trajectories efficiently, it is useful to use initial conditions in which all the particles are placed on the fixed TST dividing surface ($x = 0$) at $t = 0$. That choice guarantees that the trajectories will cross the surface at least once, but it does not prevent them from recrossing it. Consistent with the Boltzmann weighting in the flux-flux-correlation function [196, 197, 44], the initial conditions are distributed along the stable transverse direction y and the velocities according to the probability density function,

$$f(x, y, v_x, v_y) = (2\pi\, k_{\mathrm{B}}T)^{3/2}\, \omega_y\, \exp\left\{-(\omega_y^2 y^2 + v_x^2 + v_y^2)/2\, k_{\mathrm{B}}T\right\}\, \delta(x)\ . \tag{48}$$

This choice defines the barrier ensemble. The integration of the Boltzmann-weighted flux of these states gives the TST estimate of the numerator of the rate expression. If all these states were reactive and never recrossed (returned to) the fixed TST dividing surface, this estimate would be exact. The questions to be resolved below concern the deviation of the true dynamics from this TST estimate. These question will be investigated for both harmonic and anharmonic barriers. In all cases, the initial conditions will be sampled from the same barrier ensemble (48).

A stochastic trajectory is determined not only by its initial condition, but also by the specific instance of the fluctuating force that is acting upon it. In a full-fledged rate calculation, for example, an average has to be taken over both the initial conditions and the noise. The focus of this chapter, however, is the information that can be obtained about the microscopic reaction dynamics using the moving TS surface. For simplicity, a particular instance of the noise has therefore been used to illustrate most of the results. Nevertheless, the calculations presented here were repeated for several such noise sequences always leading to the same qualitative conclusions and thereby confirming that the results shown here are indeed typical. (These are not shown here for the sake of brevity.) While averages over the noise will tend to wipe out much of this microscopic detail, it is instructive to confirm the convergence of the identification of trajectories in calculating averages. In what follows, the average of the forward and backward reaction probability will be used to illustrate the convergence and degree of accuracy achievable using the moving TS surface to identify the reactive trajectories.

5.4 Analytic Results

Although anharmonicities have to be addressed in a typical chemical system, it is helpful to begin with the harmonic limiting case because it is amenable to an analytic treatment.

On the one hand, the harmonic limit illustrates the level of microscopic detail in which the moving TST method allows one to describe the reaction mechanism. On the other hand, the analytic results derived here provide a benchmark against which the performance of the numerical calculations of Section 5.5 can be assessed.

5.4.1 Reaction Probabilities

The fate of individual trajectories in the barrier ensemble (48) can easily be determined if their initial conditions are transformed into relative coordinates. The projection onto the reactive degree of freedom is illustrated in Figure 19. Since in space-fixed coordinates the barrier ensemble is centered around $\vec{q} = \vec{v} = 0$, the distribution function in relative coordinates is peaked at the stochastic position $\Delta\vec{q} = -\vec{q}^{\ddagger}_{\alpha}(0)$, $\Delta\vec{v} = -\vec{v}^{\ddagger}_{\alpha}(0)$. It reads explicitly

$$\begin{aligned} f_{\mathrm{rel}}(\Delta x, \Delta y, \Delta v_x, \Delta v_y) &= (2\pi\, k_{\mathrm{B}}T)^{3/2}\,\omega_y \\ &\exp\left\{-(\omega_y^2(\Delta y + y^{\ddagger}_{\alpha})^2 + (\Delta v_x + v^{\ddagger}_{x\alpha})^2 + (\Delta v_y + v^{\ddagger}_{y\alpha})^2)/2\,k_{\mathrm{B}}T\right\}\,\delta(\Delta x + x^{\ddagger}_{\alpha})\ . \end{aligned} \tag{49}$$

The forward-reactive part of the ensemble is formed by those trajectories whose initial velocity Δv_x is so large that the trajectory lies above both the stable and the unstable manifold of the equilibrium point. The knowledge of the eigenvectors (45) allows one to locate these separatrices quantitatively. Reactive trajectories are thus found to be characterized by the condition

$$\Delta v_x > \Delta v_{x,\mathrm{min}} := \begin{cases} -x^{\ddagger}_{\alpha}\epsilon_{\mathrm{s}} & : \quad x^{\ddagger}_{\alpha} > 0\ , \\ -x^{\ddagger}_{\alpha}\epsilon_{\mathrm{u}} & : \quad x^{\ddagger}_{\alpha} < 0\ . \end{cases} \tag{50}$$

Therefore, the probability for a member of the barrier ensemble to be forward-reactive is given by

$$\begin{aligned} P_{\mathrm{f}} &= \int d\Delta x \int_{\Delta v_x > \Delta v_{x,\mathrm{min}}} d\Delta v_x \int d\Delta y \int d\Delta v_y\ f_{\mathrm{rel}}(\Delta x, \Delta y, \Delta v_x, \Delta v_y) \\ &= (2\pi\, k_{\mathrm{B}}T)^{-1/2} \int_{\Delta v_{x,\mathrm{min}}}^{\infty} d\Delta v_x \exp\left\{-(\Delta v_x + v^{\ddagger}_{x\alpha})^2/2\,k_{\mathrm{B}}T\right\} \\ &= \frac{1}{2}\,\mathrm{erfc}\left(\frac{\Delta v_{x,\mathrm{min}} + v^{\ddagger}_{x\alpha}}{\sqrt{2\,k_{\mathrm{B}}T}}\right)\ , \end{aligned} \tag{51}$$

which has been written in terms of the complementary error function [199]

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2)\, dt \; . \tag{52}$$

In a similar manner, backward-reactive trajectories satisfy

$$\Delta v_x < \Delta v_{x,\mathrm{max}} := \begin{cases} -x_\alpha^\ddagger \epsilon_\mathrm{u} & : \quad x_\alpha^\ddagger > 0 \; , \\ -x_\alpha^\ddagger \epsilon_\mathrm{s} & : \quad x_\alpha^\ddagger < 0 \; , \end{cases} \tag{53}$$

and their probability in the ensemble is

$$P_\mathrm{b} = \frac{1}{2} \operatorname{erfc}\left(-\frac{\Delta v_{x,\mathrm{max}} + v_{x\alpha}^\ddagger}{\sqrt{2\, k_\mathrm{B} T}} \right) \; . \tag{54}$$

5.4.2 Reaction Times

In contradistinction to a space-fixed dividing surface, the moving TS surface is crossed once and only once by each reactive trajectory. This allows us to define a unique reaction time $\Delta t^\ddagger$ for each reactive trajectory: It is the time when the trajectory crosses the dividing surface, relative to the initial time when the coordinates are specified by the barrier ensemble. If the initial conditions $\Delta x_\mathrm{u}(0)$ and $\Delta x_\mathrm{s}(0)$ in the reactive degree of freedom are prescribed, the reaction time can be calculated explicitly. The dynamics of the reactive degree of freedom is given by

$$\Delta x_\mathrm{u}(t) = \Delta x_\mathrm{u}(0) e^{\epsilon_\mathrm{u} t} \; , \qquad \Delta x_\mathrm{s}(t) = \Delta x_\mathrm{s}(0) e^{\epsilon_\mathrm{s} t} \; . \tag{55}$$

The dividing surface is characterized by the condition $\Delta x = 0$, which can be rewritten in relative coordinates as $\Delta x_\mathrm{u} = -\Delta x_\mathrm{s}$. The reaction time $\Delta t^\ddagger$ at which this condition is satisfied is easily found to be

$$\Delta t^\ddagger = \frac{1}{\epsilon_\mathrm{u} - \epsilon_\mathrm{s}} \ln \frac{-\Delta x_\mathrm{s}(0)}{\Delta x_\mathrm{u}(0)} = \frac{1}{\epsilon_\mathrm{u} - \epsilon_\mathrm{s}} \ln \frac{\Delta v_x(0) - \epsilon_\mathrm{u} \Delta x(0)}{\Delta v_x(0) - \epsilon_\mathrm{s} \Delta x(0)} \; . \tag{56}$$

It is defined for all initial conditions that are either forward- or backward-reactive. For a forward-reactive trajectory, $\Delta v_x(0) > 0$. Because $\epsilon_\mathrm{u} > 0$ and $\epsilon_\mathrm{s} < 0$, it can easily be seen from equation (56) that $\Delta t^\ddagger > 0$ if $\Delta x(0) < 0$, as it should be for trajectories that start on the reactant side of the dividing surface and are still to cross it. Similarly, a trajectory with

$\Delta x(0) > 0$ is already on the product side, and its reaction time is negative. A backward-reactive trajectory, on the other hand, has an initial velocity $\Delta v_x(0) < 0$. In this case, $\Delta t^\ddagger < 0$ if $\Delta x(0) < 0$ and $\Delta t^\ddagger > 0$ if $\Delta x(0) > 0$.

If the initial position $\Delta x(0)$ is fixed, the reaction time (56) tends to zero as $\Delta v_x(0) \to \infty$: Trajectories with large initial velocities cross the barrier fast. On the other hand, as the separatrices that bound the reactive region are approached, i.e. $\Delta v_x(0) \to \epsilon_\mathrm{u} \Delta x(0)$ if $\Delta x(0) > 0$ or $\Delta v_x(0) \to \epsilon_\mathrm{s} \Delta x(0)$ if $\Delta x(0) < 0$, trajectories keep barely enough energy to cross the barrier, and their reaction times tend to $+\infty$ or $-\infty$, respectively.

Once the reaction time is given as a function of initial conditions, the distribution for the forward- or backward-reactive part of the barrier ensemble (48) is readily obtained. In the former case, its probability distribution function is given by

$$p(\Delta t) = \frac{1}{P_\mathrm{f}} \int d\Delta x \int_{\Delta v_x > \Delta v_{x,\mathrm{min}}} d\Delta v_x \int d\Delta y \int d\Delta v_y \\ f_\mathrm{rel}(\Delta x, \Delta y, \Delta v_x, \Delta v_y)\, \delta(\Delta t - \Delta t^\ddagger(\Delta x, \Delta v_x, \Delta y, \Delta v_y)) \,. \quad (57)$$

The normalization factor $1/P_\mathrm{f}$ accounts for the fact that only the forward-reactive part of the ensemble contributes to the distribution.

The distribution function (57) can in its most convenient form be written in terms of the dimensionless scaled time $\Delta\tau^\ddagger = (\epsilon_\mathrm{u} - \epsilon_\mathrm{s})\Delta t^\ddagger$. It then reads

$$\begin{aligned} \tilde{p}(\Delta\tau^\ddagger) &= \frac{1}{\epsilon_\mathrm{u} - \epsilon_\mathrm{s}}\, p(\Delta\tau^\ddagger/(\epsilon_\mathrm{u} - \epsilon_\mathrm{s})) \\ &= \frac{|r|}{\sqrt{\pi} P_\mathrm{f}} \frac{e^{\Delta\tau^\ddagger}}{(1 - e^{\Delta\tau^\ddagger})^2} \exp\left\{ -\left(\frac{r}{1 - e^{-\Delta\tau^\ddagger}} + w \right)^2 \right\} , \end{aligned} \quad (58)$$

where

$$r = \frac{q_\alpha^\ddagger(0)\,(\epsilon_\mathrm{u} - \epsilon_\mathrm{s})}{\sqrt{2\,k_\mathrm{B} T}} , \quad (59)$$

$$w = \frac{v_\alpha^\ddagger(0) - \epsilon_\mathrm{u} q_\alpha^\ddagger(0)}{\sqrt{2\,k_\mathrm{B} T}} . \quad (60)$$

The reaction probability P_f can be written in terms of r and w as

$$P_\mathrm{f} = \begin{cases} \frac{1}{2}\,\mathrm{erfc}(r + w) & : \quad r > 0 , \\ \frac{1}{2}\,\mathrm{erfc}(w) & : \quad r < 0 . \end{cases} \quad (61)$$

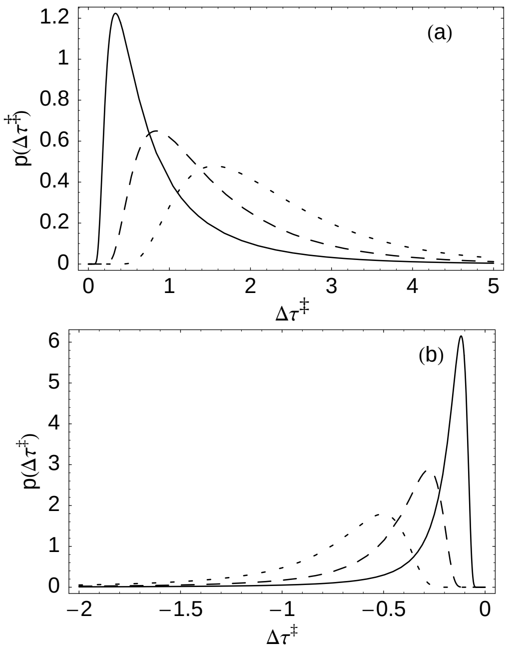

Figure 20: The distribution (58) of reaction times (a) for $w = 1$ and $r = 0.2$ (solid), $r = 0.5$ (dashed) and $r = 1$ (dotted), (b) for $w = -1$ and $r = -0.2$ (solid), $r = -0.5$ (dashed) and $r = -1$ (dotted).

The valid range of $\Delta\tau^\ddagger$ is $0 < \Delta\tau^\ddagger < \infty$ if $q_\alpha^\ddagger(0) > 0$ and $-\infty < \Delta\tau^\ddagger < 0$ if $q_\alpha^\ddagger(0) < 0$. The distribution function (58) is normalized so that its integral over that range is one. Remarkably, the distribution depends only on the two parameters r and w, even though the system dynamics and the distribution of initial conditions are determined by the five parameters ω_b, γ, T, $q_\alpha^\ddagger(0)$ and $v_\alpha^\ddagger(0)$.

In a similar manner, the distribution of reaction times can be computed for the backward-reactive part of the ensemble. The result is again given by equation (58), except that the valid range is now $-\infty < \Delta\tau^\ddagger < 0$ if $q_\alpha^\ddagger(0) > 0$ and $0 < \Delta\tau^\ddagger < \infty$ if $q_\alpha^\ddagger(0) < 0$. To obtain the proper normalization, the reaction probability P_f in the prefactor of equation (58) must be replaced by the backward-reaction probability P_b, which in terms of the scaled parameters reads

$$P_\mathrm{b} = \begin{cases} \frac{1}{2}\,\mathrm{erfc}(-w) & : \quad r > 0\,, \\ \frac{1}{2}\,\mathrm{erfc}(-r-w) & : \quad r < 0\,. \end{cases} \tag{62}$$

As can be seen from Figure 20, the reaction-time distribution (58) is highly asymmetric around its peak. The probability distribution function is flat at $\tau = 0$, where all of its derivatives are zero. For large $|\tau|$, it decays exponentially like

$$\tilde{p}(\Delta\tau^{\ddagger}) \approx \begin{cases} \frac{r\,e^{-(r+w)^2}}{\sqrt{\pi}\,P_{\mathrm{f}}}\, e^{-\Delta\tau^{\ddagger}} & : \quad q_{\alpha}^{\ddagger}(0) > 0, \quad \Delta\tau^{\ddagger} \to +\infty\ , \\ \frac{r\,e^{-w^2}}{\sqrt{\pi}\,P_{\mathrm{f}}}\, e^{\Delta\tau^{\ddagger}} & : \quad q_{\alpha}^{\ddagger}(0) < 0, \quad \Delta\tau^{\ddagger} \to -\infty\ . \end{cases} \tag{63}$$

Because the distribution is so highly asymmetric, the average reaction time will be significantly larger than the most probable reaction time that is indicated by the maximum of the distribution function.

5.5 Numerical Results

As soon as the anharmonicities of the potential in a realistic chemical system have to be taken into account, the equations of motion can no longer be solved analytically, and recourse must be taken to numerical methods. In what follows, the initial conditions, at $t = 0$, are chosen from the distribution specified in equation (48). All trajectories are evolved forward and backward in time to $t = \pm T_{\mathrm{int}}/2$ using the stochastic integration algorithm introduced by Ermak and Buckholz [139, 140]. For the backward propagation, the integration scheme was modified as described in Appendix C. In a conventional calculation of the exact rate expression, reactive trajectories are identified according to the positions they attain at the start and end of the integration interval: Trajectories that at $t = -T_{\mathrm{int}}/2$ and $t = +T_{\mathrm{int}}/2$ are located on opposite sides of the space-fixed dividing surface $x = 0$ are classified as forward- or backward-reactive; others are classified as nonreactive. This criterion, however, is only reliable if the total integration time T_{int} is sufficiently large. At short times, recrossings of the dividing surface introduce unavoidable errors.

An alternative criterion for the identification of reactive trajectories is obtained if the space-fixed dividing surface is replaced by the moving TS surface described above. In the most naive implementation, trajectories can be classified as reactive if they are on opposite sides of the moving TS surface at $t = \pm T_{\mathrm{int}}$. If the moving-TS-surface algorithm is used instead, T_{int} can be reduced by as much as a factor of 2 while still obtaining nearly accurate results. In addition, given that the moving TS surface is exactly free of recrossings in the harmonic limit and approximately so in an anharmonic potential, the integration can be stopped as soon as a trajectory crosses the moving surface: There is no need to follow the trajectory further and check for recrossings. Therefore, when the moving TS surface is used, the actual integration time will on average be much smaller than the nominal integration time T_{int}.

The reliability of this identification is illustrated below using the two-dimensional saddle point potential of equation 34 with and without anharmonicity, k. In all of the numerical calculations, the units are chosen for simplicity such that $k_{\text{B}}T = 1$. The friction is isotropic, with $\gamma = 0.2$ in these units, and selected so as to be near the turnover between the energy- and space-diffusion limited regimes. Although most of the calculations assume the same fixed noise sequence, averages of the forward and backward reaction probabilities over the noise are also shown below. In the former, the number of trajectories is fixed at $N_{\text{t}} = 15\,000$, which is large enough to make statistical errors negligible. In the single-noise calculations on the two-dimensional harmonic barrier, the transverse frequency $\omega_y = 1.5$, and the barrier frequency is set to $\omega_x = 1.0$. The latter is reduced to $\omega_x = 0.75$ for the noise-averaging and in the nonlinear cases in order to accentuate the nonlinear coupling.

5.5.1 Harmonic Systems

A typical reactive trajectory and the TS trajectory in the harmonic limit ($k = 0$), are shown in Figure 21. Clearly, the space-fixed dividing surface $x = 0$, in contrast to the moving TS surface, is crossed many times. The respective percentages of trajectories classified either as reactive and nonreactive using the fixed dividing surface are displayed as a function of integration time in Figure 22. Because all trajectories start on the dividing surface, at very short times, every trajectory is classified as either forward- or backward-reactive. Subsequent recrossings of the transition state result in transient fluctuations of the reaction probabilities that slowly approach the true, long-time values. Figure 22 also shows the percentage of trajectories that are nonreactive as well as those that cross the fixed dividing surface only once at $T_{\text{int}} = 0$. The latter comprise the majority of the reaction events, whereas the percentage contributed by reactive trajectories is comparatively small. Nevertheless, the fluctuations in the computed reaction probabilities that are caused by recrossings are considerable.

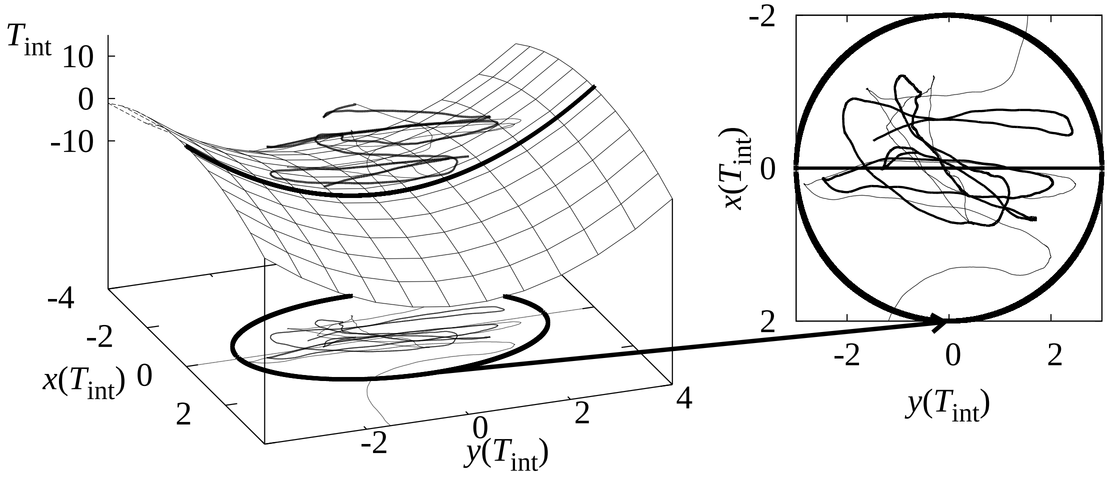

Figure 21: The evolution of a member of the ensemble and the Transition State Trajectory depicted as the gray and black lines, respectively. The underlying potential is included, and the fixed transition state $x = 0$ is highlighted by the heavy black line. The time-independent projection is shown on the base of the figure. The sample trajectory is backward-reactive since it is a reactant in the future and product in the past. As can be seen, the Transition State Trajectory remains in the vicinity of the barrier for all times.

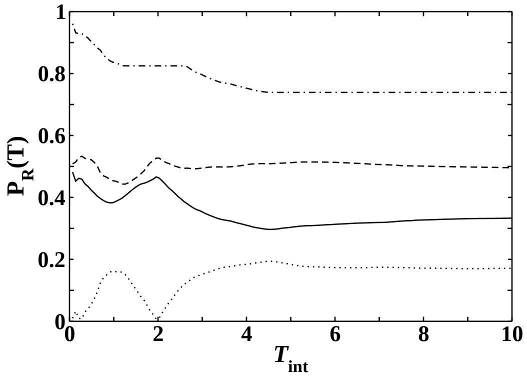

Figure 22: Reaction probabilities calculated using the fixed dividing surface displayed as a function of total integration time. The fractions of forward-reactive, backward-reactive, and nonreactive trajectories are shown as the solid, dashed, and dotted lines, respectively. The dash-dotted line represents the fraction of trajectories that cross the surface only once at $T_{\text{int}} = 0$. In these simulations, $N_{\text{t}} = 15000$ trajectories were integrated, the friction constant $\gamma = 0.2$, and the barrier frequency is $\omega_x = 1$.

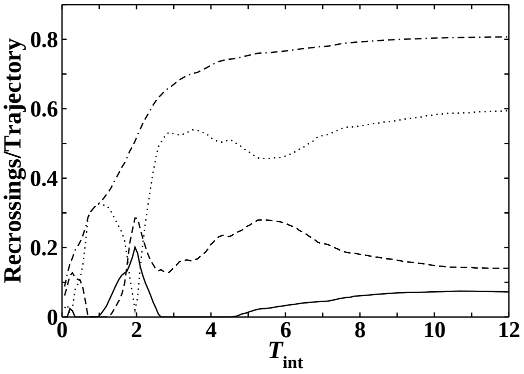

Figure 23: The number of *recrossings* (dot-dashed line) of the fixed transition state normalized by the total number of trajectories. The components of the total that resulted in a forward-reactive or backward-reactive trajectory are shown as the solid and dashed lines, respectively. Recrossings that resulted in a nonreactive trajectory are displayed as the dotted line. The simulation parameters are the same as in Figure 22.

Because recrossings are crucial to the performance of the algorithm, it is instructive to analyze them in more detail. Figure 23 shows the average number of recrossings per trajectory as a function of the total integration time. The trivial crossing of the dividing surface that all trajectories undergo at $t = 0$ is not included. The number increases monotonically as the trajectories cross and recross the transition state. Eventually, it reaches a plateau as they leave the barrier region and are lost into either the product or reactant states. In addition, Figure 23 decomposes the total number of recrossings into those recrossings that occur on trajectories that are found to be forward-reactive, backward-reactive or nonreactive at the given integration time. Because the classification of a particular trajectory can change with increasing integration time, these contributions are not monotonic. Most prominently, as the number of nonreactive trajectories decreases almost to zero at $T_{\mathrm{int}} \approx 2$ (see Figure 22), the contribution of nonreactive trajectories shows the same behavior. For large integration times, the largest contribution to recrossings stems from nonreactive trajectories, which are bound to recross the dividing surface at least once. In fact, a comparison of Figure 22 and Figure 23 reveals that nonreactive trajectories on average recross more than three times before they finally leave the barrier region. Most of the reactive trajectories, by contrast, do not recross, and their contribution to the recrossing statistics is much smaller. Asymptotically, both forward and backward reactive trajectories recross on average approximately

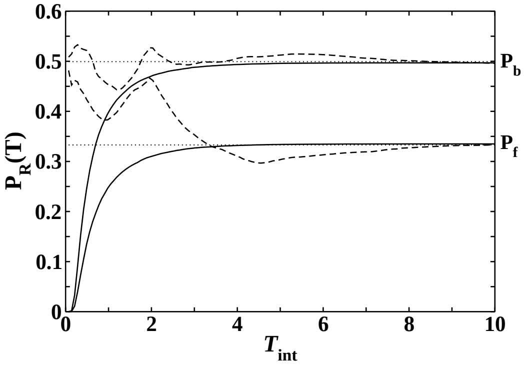

Figure 24: Reaction probabilities as a function of integration time calculated using the moving (solid curves) or the fixed (dashed curves) dividing surface. The upper set of curves represents forward-reactive probabilities, with the lower set depicting the corresponding back reactions. The dotted curves indicate the asymptotic values $P_{\text{f}} = 0.3332$ and $P_{\text{b}} = 0.4993$ calculated from equation (61) or (62), respectively. The simulation parameters are the same as in Figure 22.

0.25 times.

The dynamics is greatly simplified if the moving TS surface is used instead of the fixed one. Reaction probabilities computed using either surface are compared in Figure 24. Because the trajectories start at a distance from the moving TS surface, the corresponding rates are zero for short integration times. They then steadily increase toward the true long-time probabilities. Since the dividing surface cannot be recrossed, the asymptotic values are approached monotonically. The erratic fluctuations of the computed reaction probability that the fixed surface produces are absent if the moving TS surface is used, so that a strict lower bound for the reaction probability is obtained even for very short integration times. In quantitative terms, the moving TS surface identifies a trajectory as reactive if its reaction time $\Delta t^\ddagger$ lies within the integration interval, so that the finite-time reaction probability for a forward reaction is given by

$$P_{\text{f}}(T_{\text{int}}) = P_{\text{f}} \times \text{Prob}\left\{ |\Delta t^\ddagger| < \frac{T_{\text{int}}}{2} \right\} , \tag{64}$$

and a similar expression for the backward-reaction probability. The reaction probabilities computed from the moving TS surface are therefore determined by the distribution (58) of reaction times. The convergence toward the long-time probability is described by the long-time tail (63) of the reaction-time distribution and is exponentially fast. Indeed, Figure 24 shows that reaction probabilities computed using the moving TS surface converge much faster than those obtained from the fixed surface. Moreover, in cases such as the current problem, in which the separatrices between reactive and non-reactive trajectories are known exactly, the reaction probabilities P_f and P_b can be computed *a priori*, without having to perform any numerical simulations. The values obtained from equations (61) and (62) are also indicated in Figure 24. They agree precisely with the asymptotic probabilities obtained from the simulation. Thus, the moving TS surface can provide accelerated convergence in the rate for finite-time computations for linear problems.

The analytic reaction probabilities, equations (61) and (62), for the harmonic barrier represent the limiting values that are obtained for one instance of the noise using a large number N_t of trajectories. To obtain a macroscopically observable reaction probability, one has to average these results over a large number N_ξ of realizations of the noise. That average cannot be obtained analytically, but it can be easily calculated by a numerical quadrature. It provides a useful benchmark for the convergence of the computational schemes with respect to N_t and N_ξ. Figure 25 illustrates the forward and backward reaction probabilities, averaged over N_ξ realizations of the noise, as a function of N_ξ and for different values of N_t. The solid and dashed curves are obtained if reactive trajectories are identified through the criteria provided by the fixed and the moving TS surfaces, respectively. As expected for a symmetric barrier, forward and backward reaction probabilities converge toward the same limit. Moreover, the distributions of forward and backward reaction probabilities agree, as shown in Figure 26. For large N_t, the results in Figure 25 agree with the analytic value displayed as the dot-dashed curve in the figure's bottom panel. Therein, dotted curves are used to indicate the 95% confidence interval to further illustrate that the simulations are converging toward the correct limit as expected.

The simulation results in Figure 25 that employ the conventional criterion for identifying trajectories have been computed using the large integration time $T_\mathrm{int} = 21.5$, to illustrate the exact results within the error bars of the number average. However, it should be clear from Figure 22 that the moving-TS-surface criterion often identifies reactive trajectories in less than half this time, and once so identified a trajectory does not need to be integrated further.

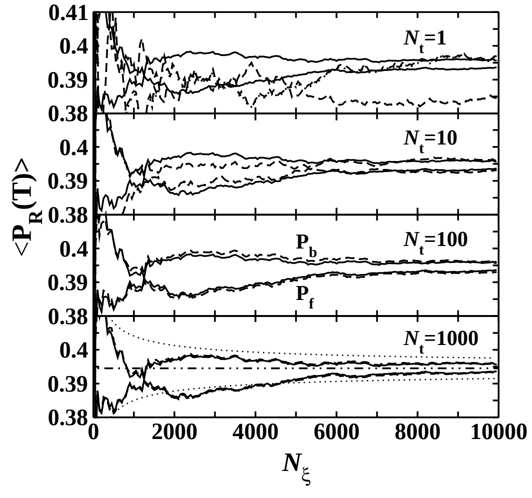

Figure 25: The reaction probabilities averaged over different instances of the noise on the harmonic potential for four different values of the number of trajectories (N_{t}) in the ensemble. The solid lines depict the results predicted by equation (61) or (62). The dashed and dotted lines are the results obtained using the respective fixed or moving dividing surfaces. In the harmonic case, these two surfaces provide the same results and are indistinguishable. For the case of $N_{\mathrm{t}} = 1000$ the light dotted lines display the 95% confidence interval with respect to the number of noise sequences sampled (N_ξ). The simulation parameters are the same as those defined in Figure 27.

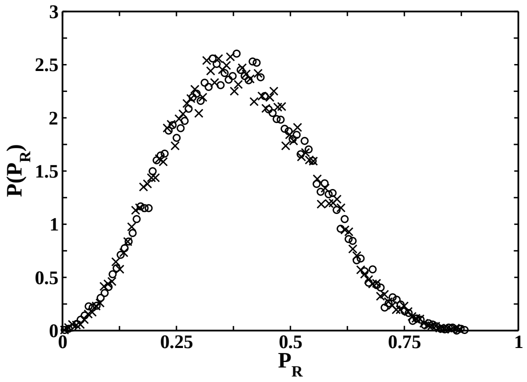

Figure 26: The distribution of reaction probabilities in the harmonic limit calculated from equation (61) or (62) from $N_\xi = 20000$ different instances of the noise. The x-symbols display the results for forward-reactive probabilities and the o-symbols are for backward-reactive. The simulation parameters are the same as those defined in Figure 27.

Given that the calculation of the moving surface itself —which amounts to the calculation of the TS trajectory— takes roughly as much computational effort as the integration of an ensemble trajectory, computational savings can thus be obtained from the use of the moving surface whenever the number N_{t} of trajectories per noise sequence is larger than 2.

5.5.2 Nonlinear Systems

The true test for the usefulness of the moving transition state lies in its ability to identify reactive trajectories beyond the linear regime. If nonlinearities are present, the relative coordinate (41) does not achieve a complete separation of the relative motion from the motion of the TS Trajectory. Therefore, the moving dividing surface will not strictly be free of recrossings. However, if the nonlinearities are weak, recrossings can be expected to be rare. In these cases, the moving dividing surface will be recrossing-free to a useful approximation. Indeed, our results indicate that its advantages over a fixed dividing surface persist well beyond the harmonic limit.

We investigate the performance of the moving dividing surface in the example of the potential (34), with the coupling constant k now taking non-zero values. The reaction probabilities for several different values of k are displayed in Figure 27. To accentuate the anharmonicity, the barrier frequency was reduced to $\omega_x = 0.75$ to allow trajectories to spend more time in the barrier region before escaping. For the transverse frequency, the value $\omega_y = 1.5$ was retained. Evidently, for sufficiently long integration times the moving transition state provides essentially the same result as the fixed dividing surface for all values of the coupling constant up to $k = 0.1$. However, the reaction probabilities converge toward the long time limit monotonically and much faster than those computed with the fixed dividing surface. Therefore, the computational advantages that the moving surface offers in the harmonic limit persist even in the presence of quite substantial nonlinearities. Eventually, of course, the use of a moving dividing surface based upon the harmonic approximation ceases to be meaningful, as can be seen for $k = 0.5$ and $k = 1$. For the specific instance of noise used in these calculations, the results obtained from the moving and fixed dividing surfaces remain in agreement for the backward-reactive trajectories, whereas a substantial

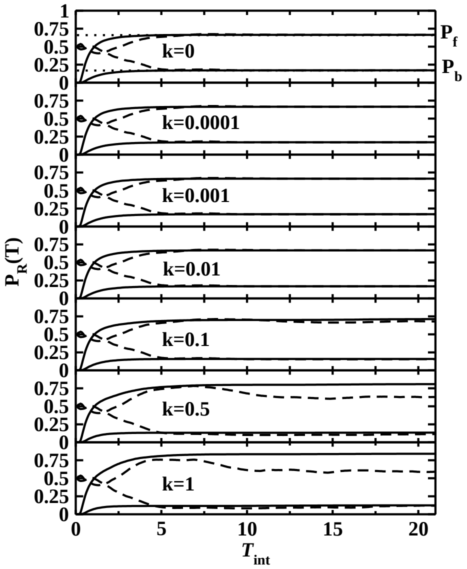

Figure 27: Reaction probabilities as a function of integration time calculated using the moving (solid line) or fixed dividing surface (dashed line) for various values of the coupling constant. The time step has been reduced to 8×10^{-6} for convergence and the barrier frequency changed to $\omega_x = 0.75$ to accentuate the nonlinearity. For the case $k = 0$, the results for the reaction probabilities as calculated from equation (61) or (62) are included as the dotted lines.

difference arises for the forward-reactive trajectories. As is to be expected of any TST scheme, in these cases the moving dividing surface overestimates the reaction probability because any trajectory that crosses the surface is assumed to be reactive, whereas the possibility of recrossings is neglected. Although not shown, a different instance of the noise does not change the trends observed in Figure 27.

As in the harmonic limit, the computational advantages of the moving dividing surface in systems with moderate nonlinearities stem from the fact that it is approximately free of recrossings. This is illustrated in Figure 28. The average number of recrossings per trajectory of the fixed transition state exhibits similar behavior for small to moderate values of the coupling constant. It approaches approximately one recrossing per trajectory in the long-time limit. In these cases, the number of recrossings of the moving dividing surface is so much smaller than the corresponding number for the fixed surface that it is not visible in the figure. At larger coupling, the number of recrossings of the fixed dividing surface does not converge to a finite long-time limit, but instead increases linearly with the integration

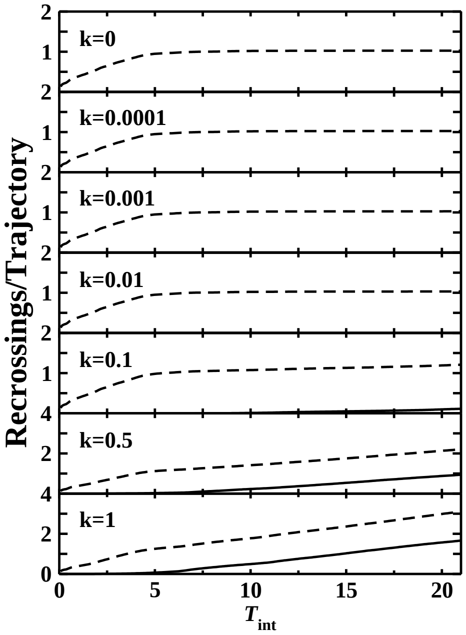

Figure 28: The average number of recrossings of the moving transition state (solid lines) and the fixed transition state (dashed line) normalized by the total number of trajectories for given values of the coupling constant. The simulation parameters are the same as those defined in Figure 27. The values for the moving transition state are too small to be seen on the same scale in the top four panels.

time. The onset of a similar behavior occurs at approximately the same value of T_{int} for the moving dividing surface as well.

This increase in the number of recrossings for both the fixed and the moving surfaces is caused by a small percentage of trajectories in the ensemble that never leave the TS region for negative times, but rather get trapped in an oscillation in the stable transverse degree of freedom y. If the value of y is sufficiently large, the reactive degree of freedom x in the potential (34) ceases to be unstable but instead behaves as a harmonic oscillator with a (possibly large) effective frequency $\tilde{\omega}_x^2 = k\, y^2 - \omega_x^2$. As a result of these fast oscillations in the reactive degree of freedom, the dividing surfaces are crossed many times. This mechanism has been confirmed by a detailed trajectory analysis, which for brevity we do not show. It is a rather peculiar feature of our model potential due to the presence of only one higher order coupling term in the potential (34). We would not expect that such aberrant behavior would arise in a typical system.

It is clear from Figure 27 that for moderately strong anharmonicities the moving transition state correctly identifies the overall number of reactive trajectories. However, that number is a macroscopic observable, and it is not immediately clear whether, on a microscopic level, individual reactive trajectories are identified correctly. The fraction of trajectories that are identified correctly by the moving transition state is displayed in Figure 29, where the “correct” identification for a given trajectory has been assumed to be given by the fixed dividing surface for a sufficiently long integration time. In the cases of weak to moderate coupling, the classification obtained from the moving dividing surface is correct for all trajectories, but, as expected, it begins to fail for coupling strengths around $k = 0.5$. The fact that the identification of the backward trajectories is poorer than that for the forward trajectories at large k is not surprising. The initial distribution —particles located at the naive fixed transition state with forward velocity— disfavors backward trajectories which must recross the fixed TS at least twice more in order to reach the appropriate boundary conditions. Nevertheless, Figure 29 confirms that the favorable behavior of the moving surface that is apparent in Figure 27 indeed reflects a correct description of the underlying microscopic dynamics.

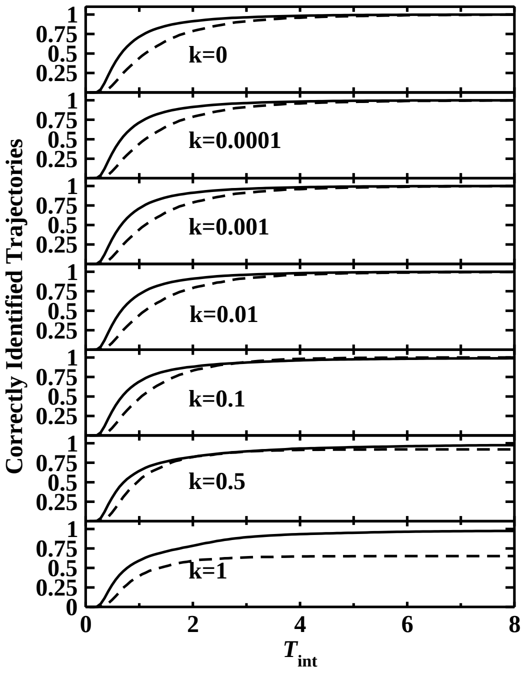

Figure 29: The fraction of correctly identified trajectories according to the moving transition state. The correct identification of a trajectory is that defined by the fixed transition state at the end of the simulation, $T_{\text{int}} = 21$. The correctly identified backward-reactive or forward-reactive trajectories are displayed as the solid and dashed lines, respectively. The simulation parameters are the same as those defined in Figure 27.

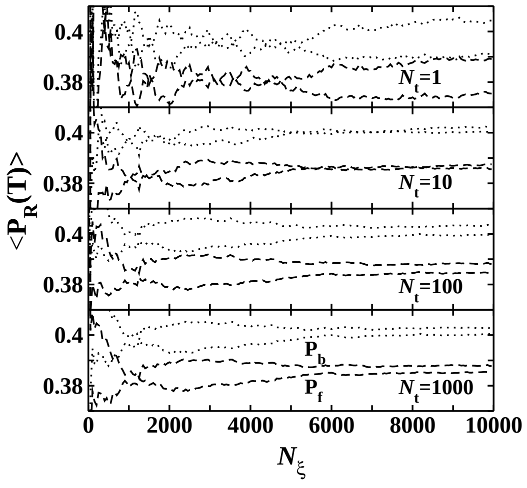

Figure 30: The reaction probabilities across an anharmonic potential (with a coupling of $k = 0.1$) are shown as a smooth function of the number of different instances of the noise (N_ξ) and a discrete function of the number of trajectories (N_t) used to represent the ensemble average. The dashed lines and dotted lines result from the use of the fixed or moving dividing surfaces in the identification of trajectories, respectively. Note that the ordinate ranges over the very narrow interval between .37 and .41, and hence the converged exact and approximate approaches are nearly equal. The simulation parameters are the same as those defined in Figure 27.

Figure 30 displays the noise-averaged reaction probabilities for the fixed and moving dividing surfaces for a coupling of $k = 0.1$. The results display the same convergence behavior in both cases, except that for the moving TS surface they are shifted to larger values by roughly 5%. This small error is due to the small percentage of trajectories that recross the moving dividing surface, as seen for a particular instance of the noise in Figure 28. Because the potential barrier described by equation 34 is symmetric even for $k \neq 0$, the average values of the forward and backward reaction probabilities P_f and P_b are equal. The simulation results converge rapidly, with respect to both N_t and N_ξ, toward their limiting value. These results demonstrate that the moving TS surface retains its reliability and its computational advantages for moderate values of the anharmonicity upon noise averaging as well as for a single instance of the noise.

5.6 Concluding Remarks

We have recently developed an analytic method for constructing a time-dependent stochastic dividing surface that is strictly free of recrossings [183, 184]. In the present chapter, it has been shown that this moving dividing surface can be used to identify reactive trajectories reliably in linear and nonlinear systems: In the harmonic limit, the moving dividing surface attached to the TS trajectory is strictly free of recrossings, while in more general (nonseparable) cases it is approximately so. The identification of reactive trajectories using the moving dividing surface has been seen in this article to be fairly accurate even in the presence of large anharmonic coupling. It can be obtained in roughly half the time that is required to confirm the nature of a trajectory by numerically evolving it to its final state.

In several of the calculations presented in this article, observables have been calculated for a particular instance of the noise while averaging over the initial conditions of the subsystem. In such restricted averages, the use of the moving surface reduces the computational cost of the calculation by a factor of two or more. A typical average of an observable, however, requires one to include multiple instances of the noise. When the average is performed using the machinery of the moving TS surface, the TS trajectory must be generated for each instance of the noise. If only one system trajectory is calculated for each noise sequence, the computational effort to calculate both the sample trajectory and the moving TS surface is about the same as calculating a single (longer) trajectory. Improved efficiency can still be obtained if one recognizes that the average should be taken by sampling several trajectories for each noise sequence. Apart from the insight into the microscopic reaction dynamics that the moving dividing surface offers, it consequently also provides computational advantages in the calculation of macroscopic observables. Moreover, the the algorithm is embarrassingly parallel when sampling across trajectories associated with a given noise sequence. (Indeed, although not discussed explicitly in the text, the codes have been parallelized across several processors with near linear scaling.)

In summary, we envision at least two approaches in which the TS-trajectory criterion for reactive trajectories will be useful in calculating reaction rates: *(i)* In harmonic (or nearly harmonic) systems, the algorithm described here provides a formally exact expression for the reaction probability given a noise sequence. This term and related averages can be used to substantially reduce the required computational time because it limits the numerical effort to a sampling of the noise. *(ii)* In arbitrary anharmonic systems, the criterion can be used to reduce the computational effort to calculate any correlation function —such as that in the reaction-rate expression— that relies on the correct identification of reactive trajectories. The rate expression and other related observables that can take advantage of the identification of reactive trajectories will be calculated in future work. As an illustration, the TS-trajectory criterion was seen in this work to converge the forward and backward reaction probabilities even in a fairly anharmonic case. Thus the central result of this chapter is: the moving dividing surface can be used reliably and efficiently to identify reactive trajectories.

CHAPTER VI

DIHEDRAL-ANGLE INFORMATION ENTROPY AS A GAUGE OF SECONDARY STRUCTURE PROPENSITY

6.1 Introduction

The number of structures in the Protein Data Bank (PDB) [200] has increased dramatically during the past decade. More than 28,000 structures had been deposited as of October 2004 when the results were first collected for this study and the count stands at at a little more than 37,000 as of July 2006. The accuracy of any new structure is of obvious importance because any error makes predictive methods more difficult to validate and creates problems for structural data-mining efforts [201, 202, 203]. As the various computational methods mature, it becomes increasingly important to derive a varied set of scores or checking functions that assess and validate protein structures. Existing and new checking functions are also vital in the area of de novo structure prediction for validation. The Ramachandran plot [204] has provided a useful framework for discerning patterns in the dihedral angle correlations and has been successfully used as a guide during structure refinement. However, it is the work of Thornton and coworkers [205, 206, 207, 208, 209] that pioneered the field of structure validation [210, 211, 212, 213, 214, 215, 216, 217] using scores based upon known statistical properties of the existing database. Although such checks are not foolproof as they rely on the working hypothesis that a new structure will "interpolate" within the known database, they at least raise the question of whether a new structure is atypical or is merely "extrapolating" outside of the database. In particular, Thornton and coworkers have proposed simple and effective ways to test the stereochemical quality of a proposed structure. Such approaches, based only on coordinates rather than on free energies or dynamical considerations, are easy to code and quick to process while still having significant merit, especially when used in conjunction with new measures.

In this study, an information-theory entropy is proposed based on the backbone dihedral angle distributions of the protein structure. It underlies an auxiliary robust checking function for evaluating the compatibility of a given protein structure with the experimentally derived structures in the PDB with respect to its dihedral angles. The twenty Ramachandran plots—*i.e.*, ϕ_i-ψ_i distributions—for each of the naturally occurring amino acids are reconstructed using all of the nonredundant experimental protein structures available in the October 2004 PDB using a 90% sequence identity cutoff. In addition, the four-hundred ψ_i-ϕ_{i+1} distributions accounting for the statistics in the two dihedral angles between specified adjacent amino acids have also been constructed and are presented. The latter distributions have been seen to contain nontrivial structure and the present results —over the existing larger database— serve to validate prior conclusions [218, 219, 220, 221]. The information-theory entropy, S, is defined in terms of the probabilities (or likelihood) of particular pairs of dihedral angles along the protein given its primary structure. A standard entropy is defined using an ideal (but likely unattainable) structure in which every angle pair, ϕ_i-ψ_i and ψ_i-ϕ_{i+1}, takes on the value with maximum probability, where the index i labels a residue along a chain. The entropy difference, ΔS, is defined relative to the standard entropy of this structure, and has been calculated for all nonredundant protein structures in the PDB. A histogram of these entropy differences leads to a nontrivial distribution. As a simple test of whether such a distribution is sensitive to differences between the theoretically and experimentally generated structures in the PDB, this distribution has been obtained for each cohort. The deviations in these distributions will be seen to emerge primarily from those theoretical structures that have been obtained using statistical information that ignores long range correlation due to, for example, secondary structural elements.

Furthermore, the distribution in ΔS can be used to define auxiliary checking functions, herein called D_1 and D_2, that characterize the degree to which the dihedral angles of a given structure are compatible with the existing database. The ΔS distribution is peaked at a nonzero value because a typical structure contains a certain degree of correlation between distant residues due to secondary structural interactions. The use of the statistical distributions in the calculation of ΔS implies that this information is included in an averaged, or mean-field like, sense. Thus D_2 can signal the existence of atypical structures whose unusual behavior is due to specific interactions between distant residues. Of course, deviations may also be due to incorrectly obtained structures, though such a determination is not available simply from the knowledge of D_2. It therefore complements the scores available in PROCHECK [205, 206] and WHAT CHECK [202] in that it includes the ψ_i-ϕ_{i+1} correlations, and it provides a simple check of the deviation from non-mean-field like structure. Hence this measure can be used to guide modeling studies and to validate experimentally derived structures, while bolstering the tools that are available to guide the formation of de novo and engineered protein structures. In fact, D_2 provides an information-rich tool to guide experiments involving the replacement or redesign of large sections of protein structure (*e.g.*, loop modeling). These new measures also complement the work of Shortle and coworkers [222, 223, 224, 225] who focus on the propensities of a given residue's dihedral angles due to the nearby structure through an energy-based scoring function rather than on the mutual probability of given residue pairs. These subtle distinctions give rise to differences in the information that the respective checking functions or scores report. Thus the central result of this work is the construction of a new checking function D_2 that complements the existing checking functions by reporting on the extent to which the propensity of the dihedral angle deviations differ in a given protein from those of the reference database.

6.2 Methods

6.2.1 ϕ_i-ψ_i and ψ_i-ϕ_{i+1} Distributions

Dihedral angle analysis [226, 227, 228, 229] of protein backbones is helpful in structure validation and modeling [205, 206, 207, 208, 209, 210, 212, 213, 214, 215, 216, 230, 231, 217, 232, 233, 234, 235, 236, 237, 238]. Conventional Ramachandran plots $P_R(\phi_i, \psi_i)$ characterize the probability distribution for angles ϕ_i and ψ_i for each R of the twenty natural amino acids, where the two dihedral angles are defined by the backbone atom sequences, C($i-1$)-N(i)-CA(i)-C(i) and N(i)-CA(i)-C(i)-N($i+1$), respectively as shown in Figure 31. An extensive analysis of the Ramachandran plots using a fairly recent edition of the PDB has been reported by Hovmöller et al. [226].

Although useful, the information contained in a Ramachandran plot is not sufficient to

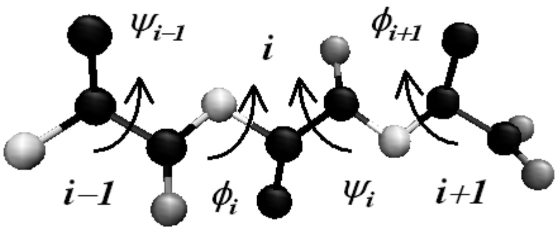

Figure 31: The backbone dihedral angles in a tripeptide ALA-ALA-ALA. Blue: nitrogen, black: carbon, red: oxygen.

construct a scoring function for high-accuracy protein structure validation. For example, flanking residues are known to affect the probability distribution in the dihedral angles of a given residue [239, 240, 241, 242, 243, 244, 245, 222, 223, 224, 225]. As previously suggested, one defines the $P_{R_i,R_{i+1}}(\psi_i, \phi_{i+1})$ distributions —in which the angles are associated with the sequential residues— to complement the information in the Ramachandran plot [218, 219, 220, 221]. Since the ψ_i-ϕ_{i+1} plot accounts for the correlation between two adjacent residues, its use in structure assessment provides a nontrivial sequence dependent measure of the likelihood that a given pair of residues will be connected by the specified dihedral angles. In principle, one could also account for the explicit correlations present between additional structural observables such as in the recent study by Esposito et al. [246] on the correlation between ψ and the angle ω describing the rotation of the peptide bond. However, only the correlation between ϕ and ψ around a residue and between bonded residues will be addressed because, as shown below, that suffices to provide a different first-order estimate of protein structure than other scores presently available.

6.2.2 Data-mining the ψ_i-ϕ_{i+1} Distributions

In order to obtain the 400 possible ψ_i-ϕ_{i+1} distributions labeled by each of the pairs of naturally occurring amino acids, a statistically representative sample of all possible proteins needs to be available. In the present work (as with other similar studies), the sublibrary of deposited structures in the PDB are assumed to be representative of the protein space once it has been systematically pruned: such as DNA, RNA and complexes of proteins with DNA or RNA, are removed. Model structures are discarded because of the unknown

Table 3: The name and number of proteins in the sublibraries used in this work are listed in the first and second columns, respectively. The peak and width of the $\Delta S^{(90)}$ distributions shown in panel c of Figure 33 and evaluated using the dihedral angle distributions from NR90 are also listed.

Sublibrary	Structures	Peak/10^{-3}	Width/10^{-4}
EXP	24,444	4.17	0.96
NR100	11,157	4.24	0.92
NR90	2,762	4.37	0.98
NR70	2,176	4.74	1.03
NR50	1,768	4.85	1.04
NR100T	644	4.44	0.79

possibility that such theoretically-derived structures may be of a different level of accuracy or representation. Additionally, structures with missing residues or containing unified atoms have been removed. (Although more aggressive pruning could have been done by discarding structures according to a more rigorous standard for its resolution, this was not done in this investigation.) After pruning the PDB subject to these criteria, the resulting library (called "EXP" throughout this work) includes a total of 24,444 experimentally derived structures.

The NR50, NR70 and NR90 sublibraries result from the intersection of the EXP library of October 2004 PDB structures with the nonredundant sequence databases posted in the PDB —as listed in the April 2005 update— at the 50%, 70% and 90% sequence identity levels, respectively [247]. The NR100 sublibrary is a subset of the EXP library in which a single arbitrarily-chosen structure is retained for each redundant sequence at 100% sequence identity. Note that, by definition, no two structures in a given database share a sequence identity greater than or *equal to* that of the database's defining percentage level. Hence, for example, the NR100 sublibrary will be smaller than the EXP library as the former includes only one structure for a given sequence. The subset, NR100T, of theoretically-derived —*viz.*, model— protein structures in the PDB at 100% sequence identity will also be investigated for confirmation of the relative level of information contained therein. The number of structures in each library is shown in Table 3.

All 400 ψ_i-ϕ_{i+1} and 20 Ramachandran plots have been generated for each of the five

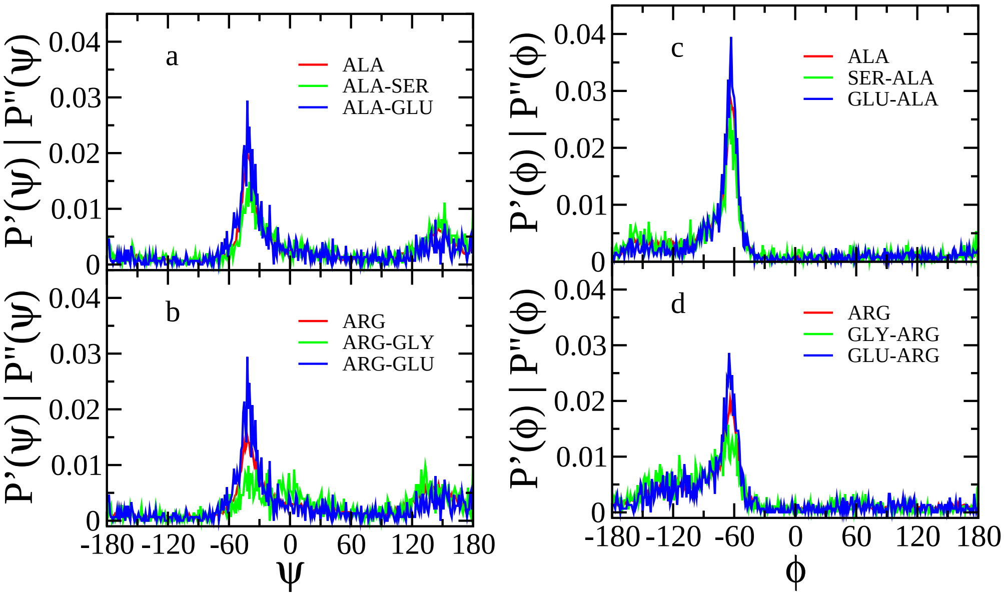

Figure 32: 1-D distributions of ψ and ϕ projected from ϕ_i-ψ_i and ψ_i-ϕ_{i+1} plots. $P'_{\text{ALA}}(\psi)$, $P''_{\text{ALA,SER}}(\psi)$, $P''_{\text{ALA,GLU}}(\psi)$ (a); $P'_{\text{ARG}}(\psi)$, $P''_{\text{ARG,GLY}}(\psi)$, $P''_{\text{ARG,GLU}}(\psi)$ (b); $P'_{\text{ALA}}(\phi)$, $P''_{\text{SER,ALA}}(\phi)$, $P''_{\text{GLU,ALA}}(\phi)$ (c); $P'_{\text{ARG}}(\phi)$, $P''_{\text{GLY,ARG}}(\phi)$, $P''_{\text{GLU,ARG}}(\phi)$ (d).

sublibraries, NR50, NR70, NR90, NR100 and EXP. Their construction is described explicitly in Appendix D. Typical 1-D distributions of the projections of the ϕ_i-ψ_i Ramachandran plots and the ψ_i-ϕ_{i+1} plots are displayed in Figure 32. (for the procedure, *cf.* Appendix D.) These results demonstrate the sequence dependence of the ψ_i-ϕ_{i+1} distribution, in accordance with the previous reports [248, 218, 220]. Importantly, the dependence of ψ_i on the second residue and ϕ_{i+1} on the first residue obviously illustrates the impact of the distant residue identity on the absolute value of the maximum probability. The effects on glycine are particularly pronounced as the peak position of the distribution changes with the distant residue identity (Figure 2b). The 420 histogramed distributions for NR90 have been saved into a single database which can, in turn, be used to calculate the dihedral-angle information entropy difference, ΔS, defined in Eq. 70 below.

6.2.3 The Dihedral-angle Information Entropy

Given a proposed protein structure for a particular primary sequence, and the distribution functions described above, one can calculate an information-theory-based entropy for the angle pairs about and between the residues of the chain. In particular, for a given structure $\vec{q}$ containing the coordinates a given protein, the dihedral angle pairs across its n residues consist of the $(n-2)$ ϕ_i-ψ_i pairs with their associated probabilities $P_{R_i}(\phi_i, \psi_i)$ at each site i for i ranging across 2 and $n-1$. Similarly, $\vec{q}$ gives rise to the $(n-1)$ ψ_i-ϕ_{i+1} pairs and associated probabilities $P_{R_i,R_{i+1}}(\psi_i, \phi_{i+1})$ between successive residues at i and $i+1$ for i ranging across 1 and $n-1$. For convenience, these two sets are interlaced into a single vector $\vec{\Upsilon}$ whose $2n-3$ entries are defined as:

$$\Upsilon_{2i-1} \equiv (\psi_i, \phi_{i+1}) \qquad \text{for } 1 \le i \le n-1 \tag{65a}$$

$$\Upsilon_{2i} \equiv (\phi_{i+1}, \psi_{i+1}) \qquad \text{for } 1 \le i \le n-2 \,, \tag{65b}$$

A Shannon entropy rooted in information theory [249] can now be rewritten as

$$S(\vec{q}) = -\sum_{k=1}^{2n-3} P_{\xi_k}(\Upsilon_k(\vec{q})) \ln P_{\xi_k}(\Upsilon_k(\vec{q})) \,, \tag{66}$$

where the argument in $\vec{\Upsilon}(\vec{q})$ specifies the angles according to the particular structure $\vec{q}$, and the residues are paired according to

$$\xi_{2i-1} \equiv (R_i, R_{i+1}) \qquad \text{for } 1 \le i \le n-1 \tag{67a}$$

$$\xi_{2i} \equiv R_{i+1} \qquad \text{for } 1 \le i \le n-2 \,, \tag{67b}$$

corresponding to the structure of $\vec{\Upsilon}$. A standard information entropy for a given structure can be defined in terms of the most probable dihedral angles for a given primary sequence,

$$S^\circ(\vec{q}) = -\sum_{k=1}^{2n-3} \bar{P}_{\xi_k} \ln \bar{P}_{\xi_k} \tag{68}$$

in which the maximal values are defined simply as

$$\bar{P}_{\xi_k}(\vec{q}) \equiv \max_{\Upsilon_k} P_{\xi_k(\vec{q})}(\Upsilon_k) \,, \tag{69}$$

and depend on $\vec{q}$ only with respect to the specification of its primary sequence, $\vec{\xi}(\vec{q})$. The averaged entropy difference for a given structure relative to the standard can be written simply as

$$\Delta S(\vec{q}) = (S^\circ(\vec{q}) - S(\vec{q}))/(2n-3) \,. \tag{70}$$

where $(2n-3)$ is the normalization factor.

Solis and Rackovsky [250, 251] defined a similar information entropy to that of Eq. (66) for protein structure prediction. However, none of their measures emphasized the use of the ψ_i-ϕ_{i+1} distributions, and the possible correlation between neighboring amino acids that such distributions may display. Meanwhile, the GOR algorithm [252, 253] uses the statistics of the multiple sequence alignment of segments of 17 or more residues in length to predict secondary structure assignments. The approach in this article is complementary to the GOR algorithm in that both recognize the need for studying multiple residue correlations: the latter emphasizes a larger segment while limiting the number of possibilities to the secondary structural motifs whereas the former —*viz.* the present approach— emphasizes segments limited to residue pairs while extending the accessible space to that of a discretization of the two-angle space with more than 5000 bins —*viz.* possible configurations.

Given the coordinates of a protein structure, the series of dihedral angles $\{\Upsilon_k\}$ can readily be computed. The probabilities entering in the sum of the structural entropy each depend on the relative probability that the measured dihedral angles are compatible with the corresponding residue(s) they connect. That is, the probabilities entering in equation (66) are $\{P_{R_i}(w_{k(i)}, v_{l(i)}), P_{R_i,R_{i+1}}(v_{l(i)}, w_{k(i+1)})\}$ where $\phi_i \in w_{k(i)}$ and $\psi_i \in v_{l(i)}$ given that $\{w_k\}$ and $\{v_l\}$ are the partitions in the angle space used to construct the histogramed distributions. This procedure, while direct, discretizes the possible results. Smoother estimates of the dihedral-angle information entropy could be obtained using standard interpolating techniques. But this is not done here because the simpler discrete approach provides estimates of the structural entropy with sufficient accuracy to test the proposed checking functions.

6.2.4 A Checking Function for Secondary Structure Propensity

Given the normalized probability distribution, $P(\Delta S)$, and a putative structure with well-defined dihedral angles, $\{(\phi_i, \psi_i), (\psi_i, \phi_{i+1})\}$, an integrated probability function for the entropy difference can be defined by merging the left and right cumulative distribution functions as

$$I(\vec{q}) = \begin{cases} \int_0^{\Delta S(\vec{q})} P(\Delta')\, d\Delta' & \text{if } \Delta S < \overline{\Delta S} \\ \int_{\Delta S(\vec{q})}^{\infty} P(\Delta')\, d\Delta' & \text{if } \Delta S \geq \overline{\Delta S} \end{cases} \tag{71}$$

where $\overline{\Delta S}$ is the median value of ΔS. The integral I will, by definition, take the value of $\frac{1}{2}$ when evaluated at the median. The deviation relative to the median can thus be characterized by

$$D_1(\vec{q}) = \begin{cases} \ln(2I) & \text{if } \Delta S(\vec{q}) < \overline{\Delta S} \\ -\ln(2I) & \text{if } \Delta S(\vec{q}) \geq \overline{\Delta S} \end{cases} \tag{72}$$

which takes the value of 0 for the median structure, and otherwise measures the distance away from the median structure in the distribution. When D_1 is negative (positive) it signals that the deviation is below (above) the median.

In order to make the D_1 checking function even more intuitive, a new checking function D_2 is defined to roughly describe the number of standard deviations away from the median structure through the expression

$$D_2(\vec{q}) = \begin{cases} \sqrt{2}\mathrm{erf}^{-1}(2I-1) & \text{if } \Delta S(\vec{q}) < \overline{\Delta S} \\ \sqrt{2}\mathrm{erf}^{-1}(1-2I) & \text{if } \Delta S(\vec{q}) \geq \overline{\Delta S} \end{cases} . \tag{73}$$

As described in Appendix D.0.5, the D_2 checking function evaluated for a Gaussian distribution with zero mean and unit standard deviation is exactly equal to the number of standard deviations away from the median structure. Thus D_2 may be interpreted as a measure of the relative likelihood for ΔS in terms of deviations from the mean. It effectively uniformizes the distribution in the sense that it maps the original distribution precisely to the normal curve. In particular, values of $|D_2|$ larger than 3 suggest that the specified structure in a group of structures whose cumulative likelihood, while possible, is less than 0.13%. To check the effectiveness of these new scores, D_1 and D_2 are calculated separately for the EXP and the NR100T libraries below.

6.3 Results and Discussion

6.3.1 Dihedral Angle Distributions

The ψ_i-ϕ_i and ϕ_i-ψ_{i+1} dihedral angle distributions for all five libraries described in Section 6.2.2 are described in detail in Appendix D. In addition to their role in the present work, they may be of use in homology-based methods for constructing proteins. For example, Srinivasan and coworkers [218, 219, 220] have used such distributions to predict backbone conformations of short polypeptides.

6.3.2 On the Choice of the Sequence Database

In order to implement the checks of Section 6.2.4, an underlying database must be selected. The EXP library would be a poor choice because it necessarily includes multiple copies of the same structure. Theoretically-derived structures should also be ignored because they may differ from the experimental database. In order to choose which of the experimental subsets of the nonredundant sublibraries —NR50, NR70, NR90 or NR100— would be optimal, it is helpful to construct the corresponding dihedral-angle information entropy and their relative properties. In particular, the distributions of $\Delta S^{(X)}$ —based on the NRX sublibrary— have been evaluated across all the structures in each of the five sublibraries: NR50, NR70, NR90, NR100, and EXP. The statistical error in $\Delta S^{(X)}$ decreases with increasing X because the size of the sublibrary increases with X. But at the same time, the bias due to redundancy is also increasing with X.

The distributions of $\Delta S^{(X)}$ are shown in Figure 33. The EXP library and NR100 sublibrary contain several sets of structures with considerable sequence identity resulting in skewed distributions regardless of the choice of the checking function. As expected, the relative small size of the sublibraries underlying the $\Delta S^{(50)}$ and $\Delta S^{(70)}$ measures leads to noisy distributions. Meanwhile, the distributions in $\Delta S^{(100)}$ appear to be broadened by the underlying redundancy in the NR100 sublibrary. The differences between the five sublibraries appears to be revealed —and perhaps converged— most sharply by panel c which displays the distributions for $\Delta S^{(90)}$. One might be tempted to choose $\Delta S^{(70)}$ instead of $\Delta S^{(90)}$ because both scores reveal that the NR90 distribution is more like that of the redundant

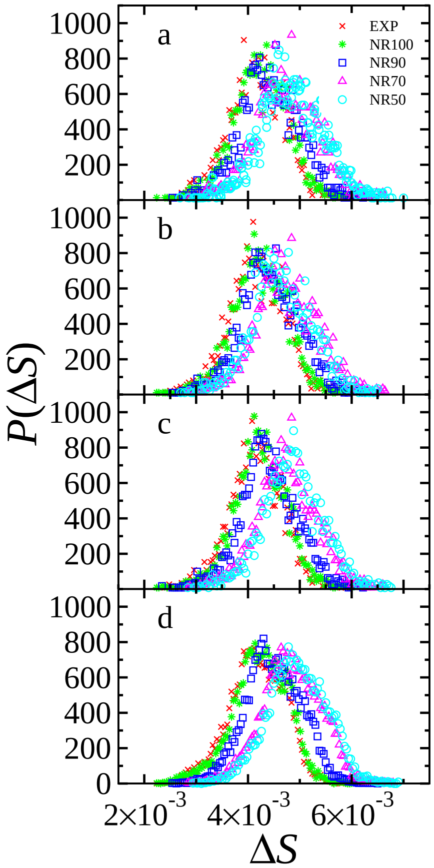

Figure 33: The distributions of $\Delta S^{(50)}$, $\Delta S^{(70)}$, $\Delta S^{(90)}$ and $\Delta S^{(100)}$ evaluated across several sublibraries are displayed in panels a, b, c and d, respectively. In each panel the information entropy difference ΔS is evaluated across the NR50, NR70, NR90 and NR100 sublibraries, and the EXP library.

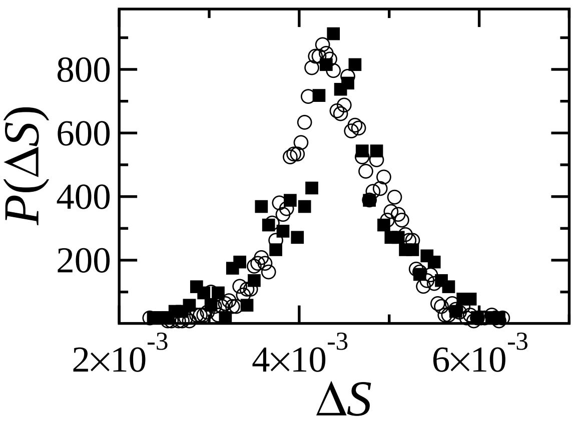

Figure 34: Distributions of ΔS evaluated across the 2,762 experimental structures in the NR90 sublibrary (circles) and 644 theoretical structures in the NR100T sublibrary (squares). Note that in order to make the results comparable, the distributions have been normalized by the bin size, *i.e.*, 4×10^{-5} and 8×10^{-4}, respectively.

libraries. However, the better statistics of $\Delta S^{(90)}$ in light of the relatively small redundancy error, and the similarity in the peak positions between NR100 and NR90 as listed in Table 3, suggests that NR90 is an optimal choice. In light of this heuristic argument, NR90 is used in the remainder of this chapter as the underlying distribution in calculating ΔS and the associated checking functions; the superscript in $\Delta S^{(90)}$ is henceforth omitted.

The distributions of ΔS for experimental and theoretical structures in NR90 and NR100T, respectively, are shown in Figure 34. The mean value and standard deviation σ of ΔS of experimental structures are 4.38×10^{-3} and 5.74×10^{-4}, respectively, indicating that roughly 71% of the total structures have a ΔS between 3.81×10^{-3} and 4.95×10^{-3}, *i.e.*, between $\langle \Delta S \rangle - \sigma$ and $\langle \Delta S \rangle + \sigma$. The mean value and standard deviation for the theoretical structures are 4.35×10^{-3} and 6.82×10^{-4}, respectively, and about 64% of the theoretical models have a ΔS within one standard deviation of the mean of the experimental models. The two distributions are surprisingly similar, particularly as the difference seen between the NR90 and NR100 distributions does not appear to persist for NR100T. The origin of this likely lies in the fact that the NR100T sublibrary does not have NR100's degree of sampling bias because the latter contains many similar single-point mutants. However, on average, fewer theoretically determined structures are within a σ of the mean and this is a notable difference between the experimental and theoretical structures. This result is likely a consequence of the fact that many theoretical structures use rule sets for their construction which do not reflect the degree of correlation between distant residues present in nature. These observations indicate the insight that ΔS provides on the relative compatibility of a given structure with respect to the experimental NR90 sublibrary of the PDB.

One possible concern here is that the only standard for inclusion of a protein within any of these libraries with respect to the accuracy of the structure lies in the fact that the reported structure provides sufficient information to obtain all of its dihedral angles. One could use more rigorous criteria employing R-factors or other self-reported position error bars. Indeed several studies that have developed checking functions have used such rigorous criteria [202, 205, 206]. However, we found that implementation of these criteria in constructing libraries nearly requires a file-by-file assessment as the requisite information is not coded uniformly through the PDB. Meanwhile our preliminary constructions of such libraries, while modifying the dihedral-angle distributions slightly, do not lead to appreciably distinct distributions in the information theory entropies or the various checking functions. Hence all the results reported here have been obtained using the simple rule for structure identification described above.

6.3.3 D_1 and D_2 Checks

The distributions of D_1 calculated using Eq. 72 across the NR90 and NR100T sublibraries are shown in Figure 35. The distributions are nearly Gaussian as suggested above. However, features seen above in Figure 34 in assessing the relative compatibility between the NR90 and NR100T sublibraries are still visible in Figure 35. The distributions in D_2 displayed in Figure 36 retain these features as well, but the uniformizing procedure outlined in Appendix D.0.5 now leads to a normal Gaussian distribution for the NR90 structures. Interestingly, the lack of correlation in some of the NR100T structures is exhibited by a shoulder on the left side of the NR100T distribution.

Although the definitions of D_1 and D_2 may appear cumbersome, their generalized forms are helpful so as to account for the fact that the probability distribution in ΔS is not

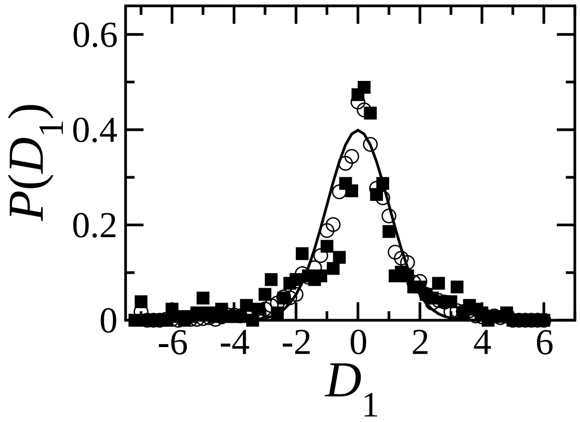

Figure 35: Distributions of D_1 evaluated across the NR90 (circles) and NR100T (squares) sublibraries. In all cases, D_1 is determined using equation 72 with $\overline{\Delta S}$ equal to the corresponding median value ($=4.38\times10^{-3}$) of the experimental structures in the NR90 sublibrary. The solid line is a Gaussian distribution with zero mean value and unit standard deviation.

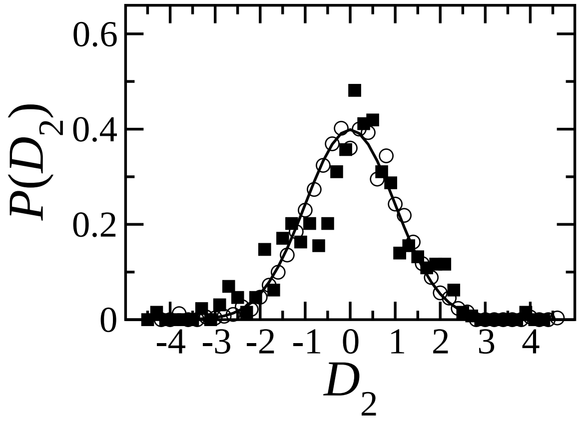

Figure 36: Distributions of D_2 evaluated across the NR90 (circles) and NR100T (squares) sublibraries. The median value of ΔS of experimental structures has been used. The solid line is the same Gaussian distribution as in Figure 35.

symmetric. If it were symmetric, then the simpler arguments at the end of the previous section using a single characteristic σ would suffice. As remarked previously (and shown explicitly in Appendix D.0.5), in the limit that the distribution in ΔS is Gaussian, the definition of D_2 reduces precisely to the number of standard deviations that a given structure differs from the median. In summary, equations 72 and 73 define equivalent new checks, D_1 and D_2, for the compatibility of the dihedral angles of a given structure with the existing PDB set of nonredundant experimental structures, although D_2 is preferred because it takes on nontrivial values even for exponentially unlikely structures.

To illustrate the values of the D_1 and D_2 checks, it is helpful to examine a few representative structures arbitrarily chosen from the PDB. The HIV envelope glycoprotein (1g9nG) [254], the p53 DNA binding domain (1tupA) [255], and the G-protein α-1 chain (1gg2A) [256] are fairly common proteins whose structures have been resolved and deposited in the PDB. The D_1 values for these structures are -0.06, -0.25 and 0.23, respectively, which alone might not seem to provide a simple score of the structural quality. However, the D_2 values are -0.08, -0.32 and 0.33. These values are easily interpreted as they indicate that all three structures are within one standard deviation of the PDB database. That is, their dihedral angles with respect to correlation about a residue and between residues are typical of the structures in the NR90 sublibrary. But recall that their information entropy is consequently greater than their corresponding standard entropies. Thus, they evidently exhibit propensities for secondary structural interactions that are typical of the structures in the NR90 sublibrary.

Alternatively, the D_2 check can be used to identify protein structures whose angles are atypical with respect to the distribution of correlated angles in the PDB. Such atypical structures are not necessarily incorrect structures. Indeed, when D_2 is large and negative, the structures could be correct but for whatever reason contain dihedral angles in the most probable positions independent of the sequence beyond their nearest neighbors. Alternatively when D_2 is large and positive, particularly strong correlations of distant residues may give rise to angles that adopt low probability configurations. Though correct structures exist that satisfy such limits, they are still atypical relative to the distribution because, as shown in Figure 34, most of the experimental structures in the NR90 sublibrary have a structural entropy difference near the mean, $\overline{\Delta S}$. This raises the intriguing possibility that D_2 can be used to highlight atypical regions in proteins that are atypical due to some functional constraint. These regions could arise for reasons related to active site architectures or regions critical to forming protein-protein interactions. Hence the D_2 measure may serve a role in highlighting regions of interest when structures of unknown function or physiological role are solved as part of ongoing high throughput structural proteomics efforts. Long range interactions through a protein structure are of course important to understanding catalysis, concerted movements, and even when seeking to understand the evolutionary history of proteins within a conserved family of proteins. Thus D_2 can highlight these potential regions within a structure too.

6.3.4 The Role of D_2 in Checking Theoretical Structures

All structures in the NR90 and NR100T sublibraries whose value of $|D_2|$ is larger than or equal to 3 are listed in Table 4. The number of such structures is 17 (0.6%) and 11 (1.7%) for the experimentally and theoretically derived structures, respectively. The structures in the larger EXP and model protein libraries have also been assessed according to the D_2 check. It was found that 264 (1.1%) and 66 (6.7%) structures are atypical out of the 24444 experimental and 981 theoretical structures available, respectively. The fact that in these sublibraries, the theoretical structures are much more likely to be atypical than the experimental structures is a possible indicator that the former is somehow different to naturally occurring structures. More importantly, the primary difference manifests as a shoulder in the distributions in the negative D_2 region. This is the region that signals structures that are near to the structures with standard entropy. Thus the dihedral angles deviate little from the most likely angles, indicating that they have not been altered by secondary interactions. It should come as no surprise that some fraction of the theoretically derived structures contain dihedral angles that lack such information. However, the important result here is that D_2 is a reporter of such propensities.

Table 4: The atypical structures —namely those structures whose $|D_2|$ value is greater than or equal to 3.0— are listed according to their PDB ID —augmented by the chain ID. The corresponding D_1 and D_2 values are also provided. Note that there are 17 experimental (left) and 11 theoretical (right) structures. Perhaps noteworthy, is the fact that the sequence identity between 1jekA and 1jekB is only 17.6%, and hence can lead to rather different values of these measures.

Experimental			Theoretical		
PDB ID	D_1	D_2	PDB ID	D_1	D_2
1a2xB	7.0	3.9	1clgA	-7.9	-4.4
1a92A	-7.0	-3.9	1l1uT	7.0	3.9
1bb1B	7.9	4.4	1lh8A	-6.1	-3.1
1czqA	-6.7	-3.3	1llkA	-6.7	-3.3
1g6uA	-7.0	-3.9	1lp0A	-6.7	-3.3
1jekA	-7.0	-3.9	1m5gT	7.0	3.9
1jekB	7.0	3.9	1n1rA	-6.1	-3.1
1jrjA	-6.7	-3.3	1opvA	-6.1	-3.1
1l2pA	-7.0	-3.9	1sewA	-7.9	-4.4
1l2yA	-6.4	-3.1	1sr1	-6.7	-3.3
1motA	-6.7	-3.3	2clgA	-6.0	-3.0
1mz9A	-6.1	-3.1			
1n7sA	-7.0	-3.9			
1nyjA	-7.0	-3.3			
1pd7B	7.9	4.4			
1qr9A	-7.0	-3.9			
1sb0B	7.0	3.9			

Table 5: The values of D_2 for series of template structures (TS) and the corresponding model structures (MS) derived from them are shown for the 1fdx target. The sequence identity (SI), and the RMSD between the MS and target are also provided. Note that the use of the target as the TS results in a different MS than the target as indicated by a nonzero value in their RMSD.

Template	SI(%)	D_2(TS)	D_2(MS)	RMSD(Å)
1fdx	100.0	-0.73	-0.03	0.26
1fdn	66.7	-1.05	-0.08	0.69
5fd1	42.6	-0.48	-0.04	1.67
1fxd	35.2	-0.34	-0.04	5.27
2fxb	20.4	-0.12	-0.06	8.15

This can be further illustrated through a study of the D_2 check on a series of structures constructed by homology modeling. The success of the homology modeling package, MODELLER [257, 235], has previously been shown using several template structures (TS) to construct model structures (MS) for the protein with PDB ID, 1fdx. This study has been reproduced here with the additional construction of the model structure based on the known 1fdx target structure as a template. The sequence identities (SI) and RMSD's are shown in Table 5, and illustrate the previously reported success. Namely, the greater the sequence identity between primary structures of the TS and target, the smaller the RMSD between the MS and the target. Although it should be noted that the RMSD is not zero even when the target structure is used as the TS. As also reported in Table 5, the D_2 check of the target structure ($= -0.73$) is far from zero, as is the value of this checking function for most of the TS's. However, the D_2 checks of all five predicted MS's are nearly zero, and all are evidently different to that of the corresponding TS and to that of the target structure. The D_2 check does not differentiate between these 5 MS's in terms of their relative fidelity to the target structure. Other scores or checking functions are needed for (and indeed some satisfy) this property. However, the consistently zero value in the D_2 checks of the MS's illustrates the fact that structures predicted by MODELLER, while often containing high fidelity to the target structure, leave out some property that would make them atypical of the PDB in the sense that is measured by the D_2 check. This property is the long-range

Table 6: The D_2 check, G-factor and Z-score values for ten different protein structures available in the PDB are listed.

PDBID	D_2	G-factor	Z-score
1tupA	-0.32	-0.02	0.41
1g9nG	-0.08	-0.14	-4.47
1gg2A	0.33	0.07	-3.56
1stn	-0.8	0.06	-0.04
1jekA	-3.9	0.75	4.23
1jekB	3.9	0.67	5.41
1n7sA	-3.9	0.62	3.77
1l1uT	3.9	0.05	-6.53
1lp0A	-3.3	0.12	3.36
1lh8A	-3.1	0.31	4.09

correlation in the dihedral angles between non-neighbor residues. While perhaps not surprising that MODELLER removes this propensity, it is nevertheless useful that D_2 check provides a quick verification of this removal and it evidently provides an independent check for what could be done to expand the functionality of programs such as MODELLER.

6.3.5 D_2 and Other Checking Functions

A comparison between D_2 check, the torsion angle G-factor in PROCHECK [205, 206] and the Ramachandran Z-score of WHAT_CHECK [202] has been made for several example structures in the PDB. The torsion angle G-factors is a log-odds score of the observed distributions of the ϕ_i-ψ_i combination. A low G-factor often indicates an unusual structure [205, 206]. The Ramachandran Z-score is the the number of standard deviations that the score deviates from the expected value. It shows how "normal" the ϕ_i-ψ_i angles in a protein structure are. Z-scores above 4.0 and below -4.0 are very uncommon [202]. The results are shown in Table 6 for six experimental structures —1tupA, 1g9nG, 1gg2A, 1stn, 1jekA, 1jekB and 1n7sA— and three theoretical structures —1lluT, 1lp0 and 1lh8A— which have been chosen because they provide a range of D_2 values. Except for two structures, 1g9nG and 1gg2A, the Z-scores are compatible with the D_2 values in terms of the assessment that the structures are typical or not typical. However, most of the structures contain G-factors that are not compatible with their D_2 checks in terms of this assessment. (Note that in order to run PROCHECK, a resolution for a structure must be specified. Although this is readily available for experimental structures, it is evidently not available for the theoretical structures. Nevertheless, the theoretical structures where run with varying resolutions —2.0, 2.5 and 3.0 Å— all resulting in the same values for the G-factors.) In summary, D_2 check differs from PROCHECK and WHAT_CHECK in their assessments of these protein structures, and evidently provides distinct information about the structures. In particular, as seen above, the use of the dihedral-angle correlation between neighboring residues in the D_2 checking function allows one to obtain a signal of the presence for propensities between residues beyond the nearest neighbor. It thereby complements the information from PROCHECK and WHAT_CHECK.

6.4 D_2***Check Web Server***

The D_2 check function has recently been incorporated into a web-based server so as to readily allow its use in the analysis of existing PDB structures or new structures at the residue and protein scales. The server also provides related information such as the Ramachandran plot and ψ_i-ϕ_{i+1} plot of a protein structure. The values of D_2Check at the residue scale can easily be summarized using a novel color strip [258] that is also generated by the server.

All of the server-side interface and file handling software has been written in perl, php and/or html with server-side includes. The user provides structural information either through a direct upload of a file that follows the PDB file protocol, or by inputing the PDB ID of the protein of interest. The default output includes the protein's Ramachandran plot, its ψ_i-ϕ_{i+1} plot, the D_2 check value for the whole protein, and a color strip summarizing the D_2 check values at the residue scale in a compact form. Several options are available to the user: one can request analysis of a particular chain within the structure, elect to receive the source files for the ϕ_i-ψ_i distribution, and/or elect to receive the source file for the D_2Check values at the residue scale. This data is provided (once available) by way an e-mail alert message containing an html link to a page with a randomized address. (The results are deleted within 7 days on the server and are unlikely to be found by a webbot. Nevertheless, in future versions, these pages will use security protocols to further ensure the

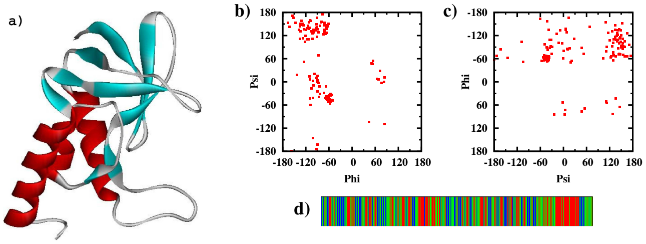

Figure 37: Given a protein structure as illustrated in panel *(a)* for the 1STN protein, the D_2Check server constructs the the Ramachandran plot (panel *(b)*), the $\psi_i - \phi_{i+1}$ plot (panel *(c)*), the D_2 value for the overall protein, and the D_2 color strip (panel *(d)*).

protection of user data from other users.)

The underlying server-side engines were developed locally and perform various calculations using the structural information available in the PDB file. It is by the use of these engines that the D_2Check server obtains the ψ_i-ϕ_{i+1} and ϕ_i-ψ_i distributions for a given structure, references the library of distributions precomputed from the PDB, calculates a structure entropy of a given protein chain, obtains the information-theory entropies, and produces the D_1 and D_2 scores [258].

The D_2Check server can be used to obtain the D_2 check values at the residue scale —when analyzing the dihedral angles between and about residues— and at the protein scale —when averaging over the entire structure. In the former case, it is convenient to provide a simple and compact visualization of the $2n - 3$ values. To this end, a color strip is produced by the server that uses a succession of colors to represent the values along the protein chain, starting with the N-terminus. Typical graphical outputs of the server are illustrated in Figure 37, for the staphylococcal nuclease (1STN) structure of [259]. Future upgrades will allow the user to make requests for multiple file (or multiple chain) processing and the systematic detailed comparison of two different sequences, including a difference strip of the corresponding D_2Check values at the residue scale.

6.5 Conclusion

A dihedral-angle information entropy describing how a particular model protein is similar to naturally occurring proteins has been discussed in this work. Based on this entropy, new checking functions, D_1 and D_2, are proposed as a check of the likelihood of the compatibility of the dihedral angles of a given structure to the experimental structures in the PDB. The results for both experimentally and theoretically derived structures in the PDB indicate that this method is simple and effective.

Generally speaking, the D_1 and D_2 checks signal the propensity for a protein to contain secondary structural interactions in comparison with the PDB. The overall structures found to be atypical by these checking functions may be classified as either: *(i)* weakly correlated (or mean-field like) in the sense that residues beyond the nearest residue do not affect the dihedral angles, *(ii)* strongly correlated in the sense that distant residues lead to large deviations in the dihedral angles away from the typical values, or *(iii)* incorrect in the sense that some of the angles may have been incorrectly assigned. In particular, large negative values of D_2 check indicate structures that are perhaps too likely, while large positive values indicate structures that are perhaps too unlikely in comparison with the the typical structures of the PDB database. The use of D_2 check at the residue level has been developed and will be discussed separately [258]. D_2 check is complementary to existing scoring functions used in assessing structure predictions but provides a different form of stereochemical information. For example, it can be used in concert with other functions to identify important or unusual parts of a structure.

One criticism that could be levied against this work —and indeed against many bioinformatic tools based on a reference set— centers on the question of whether the chosen reference sublibrary of the PDB is representative of the protein universe. The recent work of Zhang et al. [260] suggests that the diversity of single-domain structures available in the PDB database is indeed representative of the protein universe. But there may be a danger that the distribution of such structures is skewed in some way. In order to reduce the presence of such biasing, the reference sublibrary selected in this work excluded structures that had over 90% sequence redundancy. Meanwhile, the statistical information available from the current size of the database was sufficient only for bins with 5° windows. While both the coverage of the protein space and the accuracy of the distributions appear to be sufficient in the treatment performed here, one would expect that both would improve in the future as the PDB grows.

One additional result of this work is the confirmation that the ψ_i-ϕ_{i+1} plots contain correlation between dihedral angles of a given residue and the identity of the neighboring residue. This result validates previous observations [248, 218, 219, 220, 242, 244, 245]. It is seemingly in contradiction of the Flory isolated-pair hypothesis [261] in which it was assumed that the ϕ_i-ψ_i distribution of each residue in a protein backbone is independent of the neighbors' identities. However, the differences found here are sufficiently small that violations of the isolated-pair hypothesis are subtle. For this same reason, it is not surprising that Brooks and coworkers [262] found that the isolated-pair hypothesis holds very well upon averaging over the ensemble to obtain conformational entropies.

In summary, this work serves to increase the awareness of the effect of nearest neighbor frequency on the pairwise dihedral distributions and introduces a useful series of checking functions that can be used to interpret both experimental and theoretical protein structures.

CHAPTER VII

CONCLUSIONS

Current research efforts in surface science have largely been driven by experimental methods capable of creating and probing structures at ever smaller scales. Soon it should be possible to perform detailed classical simulations of systems that are of the same macroscopic size as those that can be created in the laboratory. Unfortunately, while the length scales obtainable by theory and experiment may become comparable, a large disparity in the accessible time scales will still exist. That is, without major algorithmic and theoretical advances, standard molecular dynamics will not be able to reach physically-relevant macroscopic time scales in the near future. Coarse-graining on some level will continue to be necessary. The use of projective (stochastic) models —e.g. the Langevin model— in a rigorously defined approach allows for a consistent treatment of the remaining modes whose detailed dynamics are of interest.

To demonstrate the utility of such an approach the costly molecular dynamics simulations of surface diffusion introduced in Chapter 2 provide a general prescription for extracting the relevant parameters required in coarse-grained Langevin models. This in turn, allows one to efficiently explore hypothesized mechanisms and interesting phenomenon associated with surface diffusion (cf. Chapter 3) while maintaining a physically realistic description of the system. Perhaps more importantly, the coarse-grained representation allows one to extend the range of parameters that can be explored when brute force dynamics simulations become impractical. This applies, in particular, to the low temperature regime where the rate of the diffusion process becomes too slow to follow in a reasonable amount of simulation time. However, the coarse-grained description is less sensitive to this fact and hence may prove to be a powerful method for the study of surface diffusion. A final feature gleaned from Chapter 2 confirms the adequacy of the simple phenomenological models often employed by practitioners of Langevin simulations of surface phenomenon. Periodic potentials of mean force with barrier heights related to the activation energy and dynamics that lie in the underdamped regime is precisely what have been found from the rigorous coarse-graining.

Based on these observations, the studies of diffusion over fluctuating barriers in extended systems discussed in Chapters 3 and 4 were readily performed in the coarse-grained representation with minimal computational effort. The goal of attaining atomic-level control over the motion of adsorbates on surfaces discussed therein remains an active area of research due to its potential impact on a wide range of fields. By taking advantage of the large differences in diffusion rates induced by the resonant activation, this may offer an alternative route to gaining such control. While this possibility has been confirmed in Langevin systems [137] one of the key features of this hypothesis remains to be addressed. Can this phenomenon be observed in molecular simulations of surface diffusion and if so, what is the most desirable mechanism to give rise to such behavior?

The crux of the proposed mechanism for inducing the resonant activation is to control the relaxation time of the barrier fluctuations, and this change should be directly manifested in the corresponding force autocorrelation function. Unfortunately, as seen in Figure 6, the relaxation rate of metal atoms on metal surfaces is incredibly fast, and any method for modifying the fluctuations must occur on a similar time scale. Most likely, no physically realizable approach will give rise to the desired phenomenon in this system. However there are several proposed mechanisms for increasing the relaxation time that remain to be tested. One possible method for affecting such a change is to alter the composition of the underlying subsurface layers, necessarily leading to a modification of the associated potential and forces. Alternatively, one may introduce a variety of adsorbed molecular species that interact with different modes of the surface hence giving rise to different correlations. Regardless, once a mechanism for modifying the memory time of the barriers is determined, the Langevin simulations of Chapters 3 and 4 indicate that it can presumably be utilized in a straightforward manner to control the diffusive properties of the adsorbed species.

It is questions of this type that have been the motivation for much of the methods development described here. In general terms, the ability to take advantage of the properties of complex nonequilibrium environments in order to modify or control a desired observable has played a central role in the early chapters of this work and may have a profound impact on the future design and construction of new materials. Unfortunately the tools required for such studies are not implemented in any standard simulation packages and hence each respective chapter required a sizable amount of code development. Furthermore in order to begin to address these topics, one must first gain an understanding of the system subject to these conditions, and the fundamental knowledge gained from such a study along with the methods used to obtain it are interesting and useful to the community regardless of the outcome. The tools developed in these chapters readily allow one to address a variety of systems whether it be a thin polymer film adsorbed on a surface, the complex interactions of many small molecules on a surface, or perhaps even the impact of external fields on the diffusion processes (cf. Figure 1). While such features can in some instances be explored with brute force simulation techniques, as the system becomes larger and more complex, it becomes all the more advantageous to possess alternative approaches for their simulation —such as the Langevin description— in terms of convenience, elegance and understanding.

Moreover, in some instances coarse-grained descriptions are the only option available. As in the case of polymer systems adsorbed on surfaces mentioned above, atomistic simulations involving both the dynamics of the substrate as well as the adsorbed polymeric moieties quickly become prohibitively expensive. If one is interested for example, in the the self-assembly properties of the polymers or electron transport within the film, the dynamics of the substrate are of little concern to the process at hand but are nevertheless still necessary for an accurate description of the system. The ability to rigorously subsume the surface dynamics within a Langevin description allows one to recover some of the computational cost of the simulation, while retaining an adequate representation of the process.

In addition to the simulations of surface diffusion discussed in the introductory chapters, several other interesting topics have been explored including the nature of the transition state in a noisy environment and the construction of tools for protein structure validation. The development of a time-dependent transition state that is strictly free of recrossings in harmonic systems and its ensuing implementation and validation were discussed in Chapter 5. The moving dividing surface that is attached to the transition state trajectory can serve as an alternative to the standard fixed transition state and it plays an analogous role in a noisy environment. Furthermore, the moving surface also possesses several advantageous numerical properties such as improved convergence of the desired observables and a reduced number of recrossings in anharmonic cases. The corresponding development of a rate formalism based on the transition state trajectory is currently in progress.

The rate is of primary importance for almost every chemical process and several theoretical estimates for this quantity are available. However, despite several decades of development they still suffer from a well-known deficiency in that all quickly become unreliable in anharmonic systems, leaving only recourse to costly numerical simulations in this situation. Practically speaking no chemically relevant molecular system is harmonic and this fact leads directly to the main benefit of the moving transition state. The ability to still provide a transition state that is approximately free of recrossings in these cases is a major advance when coupled with its improved convergence properties. While only simple prototypical reactive systems were studied in Chapter 5, in real systems such as those of Chapter 2 the numerical effort required to obtained a converged rate is highly nontrivial. The moving transition state has the ability to mitigate a substantial amount of this cost.

A recent extension of the transition state trajectory to Hamiltonian systems may provide the most straightforward method of applying this approach to molecular systems, although the corresponding Langevin formalism would function equally as well provided the projection to the coarse-grained representation has been performed. In molecular systems, there are often two problems associated with practical implementations of rate theory calculations. The first is related to the well-known deficiencies of the chosen dividing surface but the second is more troubling and remains in general an unresolved issue. In highly dimensional systems such as for example protein folding dynamics, it is often difficult to even define a reaction coordinate for the process, much less extract the relevant parameters for a coarse-grained representation. However, provided this requirement is fulfilled and the reaction coordinate in known —which is indeed the case for the surface dynamics simulations of Chapter 2, most simple organic reactions and isomerizations, as well as numerous other systems of chemical relevance— then only an analysis of the bath dynamics is needed to complete the construction of the moving dividing surface. The rate calculation would then proceed in the same manner as any other, except the moving transition state would be used as the reactivity criterion instead of the standard fixed surface, and hence little new machinery would be needed.

Finally the development of a checking function for protein structure validation outlined in Chapter 6 provides an additional tool for those involved with protein structure design and modeling. This challenging task often takes advantage of several independent checks to gauge the quality of a given proposed structure. Chapter 6 provides such a metric based on information theory that characterizes the compatibility of a particular model protein's dihedral angles to those in the database of naturally occurring proteins. Additionally, to facilitate its use a convenient web interface has been created providing the community with open access to these tools.

The results briefly discussed in Section 6.4 allude to a potentially very useful feature of this checking function. If a D_2 value is assigned at the residue level, instead of averaging over the entire protein as was largely the focus of Chapter 6, then it becomes immediately obvious where atypical segments of the protein are located. Furthermore, within this framework it becomes trivial to determine the areas that are effected for example by a point mutation. In this case the difference of the D_2 values evaluated at the residue level along the native and mutant protein structures displays the changes that occur in the immediate vicinity of the mutation due to modifications of the local structure as expected. But perhaps surprisingly, changes at distant residues also appear which may be spatially close to the said mutation although very far away along the sequence. This observation may have wide applicability in the community as a particularly simple indicator of the long-range correlations in proteins.

APPENDIX A

PROPERTIES OF THE STOCHASTIC BARRIERS

A.0.1 Sinusoidal Potentials

The two-dimensional potential presented here is the same as that used in previous work on the diffusive motion on coupled and uncoupled stationary surfaces [91]. Although the form in equation (14) may at first glance appear coupled, a simple point transformation, rotating the coordinate frame by $\pi/4$, can be used to define the new coordinates, $u = x + y$ and $v = x - y$. This readily reduces equation (14) to

$$U(x, y; t) = (2E_{\mathrm{b}} + \eta(t)) \left[2 - \cos(v) - \cos(u)\right] \ , \tag{74}$$

which is equivalent to the *nonstochastic* decoupled potential used in reference [91] when $\eta = 0$.

A.0.2 Merged Harmonic Oscillators

The barrier height for the MHO potentials is not given by a simple additive stochastic variable as with the sinusoidal potentials. As a consequence, the barrier height distribution, $P_{\mathrm{MHO}}(V_m^\ddagger)$, is no longer given by equation (17). However, it can easily be determined according to

$$P_{\mathrm{MHO}}(V_m^\ddagger) = P\left(\eta(V_m^\ddagger)\right) \frac{\partial \eta(V_m^\ddagger)}{\partial V_m^\ddagger} \ , \tag{75}$$

where $\eta(V_m^\ddagger)$ is related to the barrier height as before. This distribution is both sharper and slightly skewed in comparison to the Gaussian distribution in η. Therefore, the stochastic sinusoidal potentials parameterized by η with a variance of 0.2 and 0.6 correspond approximately to stochastic MHO barriers with a variance of 0.05 and 0.22, respectively. (The correspondence is due to the sharpening in the distribution, but it is approximate because the former is symmetric while the latter is slightly skewed.) Although it is possible to increase the variance in the random variable to yield stochastic MHO barrier heights with the large variances (*e.g.*, 0.2 and 0.6) initially investigated using sinusoidal potentials in

Section 3.3, it is not desirable because it leads to a significantly non-Gaussian distribution in the barrier heights. Moreover, this wide η distribution contains a significant probability for inverse barriers, *vis-a-vis* wells, in which $k_m^\ddagger < 0$. (Note that when this probability is vanishingly small as in most of the cases in this work, the dynamics are not significantly affected.) Consequently, the parameters for the sinusoidal potentials were chosen to coincide with those of the stochastic MHO.

APPENDIX B

SPACE-DEPENDENT FRICTION FOR MERGED HARMONIC OSCILLATOR POTENTIALS

The piecewise nature of the MHO potential results in a piecewise form for the associated SDF. Although incoherent, every barrier gives rise to the same averages, and hence the procedure needs to be carried out only over a small region defined by the closed interval, $[x_m^0, x_m^\ddagger]$. The limits of integration over this region can be determined from the expression for the connection points

$$x_m^- = -\frac{k_0\lambda}{2k_0 - 2k_m^\ddagger} + m\lambda \ , \tag{76}$$

where $k_m^\ddagger = -(k_0 + \eta(t))$. This can equivalently be expressed as

$$\eta(t) = -\frac{k_0\lambda}{2(x_m^- - m\lambda)} - 2k_0 \ . \tag{77}$$

At the top of the barrier, when $x_m^- = x_m^\ddagger,\ \eta(t) = \infty$. In the intermediate region for arbitrary x,

$$\begin{aligned} \eta(t) &= -\frac{k_0\lambda}{2(x - m\lambda)} - 2k_0 \\ &\equiv \eta^* \ . \end{aligned} \tag{78}$$

Otherwise, at the minimum when $x_m^- = x_m^0,\ \eta(t) = -k_0$.

Although it is apparent from the expression for the barrier height that the corresponding distribution is non-Gaussian, the resulting forces are Gaussian with the probability given by equation (17). The average force for a given x is simply the weighted average of the forces when x is in the respective regions, (x_m^0, x_m^-) and $(x_m^-, x_m^\ddagger)$, which correspond to η regions of $(-k_0, \eta^*)$ and (η^*, ∞). The resulting integral for the average value of $F(x;t)$ is now:

$$\langle F_{\mathrm{U}}(x;t)\rangle = \frac{\int_{-k_0}^{\eta^*} d\eta\, F(x)P(\eta) + \int_{\eta^*}^{\infty} d\eta\, F(x)P(\eta)}{\int_{-k_0}^{\infty} d\eta\, P(\eta)} \ . \tag{79}$$

Here, one must be careful in determining which portion of the force to use in the above equation. For example, when $\eta < \eta^*$, the majority of the force is due to the barrier portion of the potential, not the well component. The average can thus be expressed as

$$\langle F_{\mathrm{U}}(x;t)\rangle = -\int_{\eta^*}^{\infty} d\eta\, k_0(x - x_m^0)P'(\eta) + \int_{-k_0}^{\eta^*} d\eta\, (k_0 + \eta)(x - x_m^{\ddagger})P'(\eta) \ , \tag{80}$$

where $P'(\eta)$ is defined through the normalization condition, *i.e.*,

$$\int_{-k_0}^{\eta^*} d\eta\, P'(\eta) + \int_{\eta^*}^{\infty} d\eta\, P'(\eta) \equiv 1 \ , \tag{81}$$

which leads to the probability distribution

$$P'(\eta) = \frac{2}{\sqrt{2\pi\sigma^2}} \frac{\exp\left(-\frac{\eta^2}{2\sigma^2}\right)}{1 + \mathrm{erf}\left(\frac{k_0}{\sqrt{2\sigma^2}}\right)} \ , \tag{82}$$

where $\mathrm{erf}(x)$ is the standard error function. Use of the normalization condition reduces the average force to

$$\begin{aligned}\langle F_{\mathrm{U}}(x;t)\rangle \quad &= \quad k_0(x - x_m^{\ddagger}) - [k_0(x - x_m^0) + k_0(x - x_m^{\ddagger})] \int_{\eta^*}^{\infty} d\eta\, P'(\eta) \\ &\quad + (x - x_m^{\ddagger}) \int_{-k_0}^{\eta^*} d\eta\, \eta P'(\eta) \ .\end{aligned} \tag{83}$$

The remaining integrals are readily computed; the explicit form of the average force is

$$\begin{aligned}\langle F_{\mathrm{U}}(x;t)\rangle \quad &= \quad k_0(x - x_m^{\ddagger}) - \Big(k_0(2x - x_m^0 - x_m^{\ddagger})\Big) \left(\frac{1 - \mathrm{erf}\left(\frac{\eta^*}{\sqrt{2\sigma^2}}\right)}{1 + \mathrm{erf}\left(\frac{k_0}{\sqrt{2\sigma^2}}\right)} \right) \\ &\quad + (x - x_m^{\ddagger}) \sqrt{\frac{2\sigma^2}{\pi}} \left[\frac{\exp\left(-\frac{k_0^2}{2\sigma^2}\right) - \exp\left(-\frac{(\eta^*)^2}{2\sigma^2}\right)}{1 + \mathrm{erf}\left(\frac{k_0}{\sqrt{2\sigma^2}}\right)} \right] \ .\end{aligned} \tag{84}$$

The second quantity to be computed is the average of the square of the force, and the derivation follows that (above) of the average force. The limits of integration are the same and the Gaussian integrals can be calculated in the same manner. Again using the normalization requirement, the first integral is eliminated such that

$$\begin{aligned}\langle F_{\mathrm{U}}(x;t)^2\rangle \quad &= \quad k_0^2(x - x_m^{\ddagger})^2 + [k_0^2(x - x_m^0)^2 - k_0^2(x - x_m^{\ddagger})^2] \int_{\eta^*}^{\infty} d\eta\, P'(\eta) \\ &\quad + 2k_0(x - x_m^{\ddagger})^2 \int_{-k_0}^{\eta^*} d\eta\, \eta P'(\eta) + (x - x_m^{\ddagger})^2 \int_{-k_0}^{\eta^*} d\eta\, \eta^2 P'(\eta) \ .\end{aligned} \tag{85}$$

The first two integrals are the same as before, and the third can be obtained with little effort. The resulting mean squared force is

$$\begin{aligned}
\langle F_{\mathrm{U}}(x;t)^2\rangle \quad = \quad & k_0^2(x-x_m^{\ddagger})^2 + \left(k_0^2(x-x_m^0)^2 - k_0^2(x-x_m^{\ddagger})^2\right)\left(\frac{1-\operatorname{erf}\left(\frac{\eta^*}{\sqrt{2\sigma^2}}\right)}{1+\operatorname{erf}\left(\frac{k_0}{\sqrt{2\sigma^2}}\right)}\right) \\
& + \frac{4k_0\sigma^2}{\sqrt{2\pi\sigma^2}}(x-x_m^{\ddagger})^2\left[\frac{\exp\left(-\frac{k_0^2}{2\sigma^2}\right)-\exp\left(-\frac{(\eta^*)^2}{2\sigma^2}\right)}{\left(1+\operatorname{erf}\left(\frac{k_0}{\sqrt{2\sigma^2}}\right)\right)}\right] \\
& + \sigma^2(x-x_m^{\ddagger})^2\left[\frac{\operatorname{erf}\left(\frac{\eta^*}{\sqrt{2\sigma^2}}\right)+\operatorname{erf}\left(\frac{k_0}{\sqrt{2\sigma^2}}\right)}{1+\operatorname{erf}\left(\frac{k_0}{\sqrt{2\sigma^2}}\right)}\right] \\
& - \sqrt{\frac{2\sigma^2}{\pi}}(x-x_m^{\ddagger})^2\left[\frac{\eta^*\exp\left(-\frac{(\eta^*)^2}{2\sigma^2}\right)+k_0\exp\left(-\frac{k_0^2}{2\sigma^2}\right)}{1+\operatorname{erf}\left(\frac{k_0}{2\sigma^2}\right)}\right]\ . \qquad (86)
\end{aligned}$$

The SDF for the MHO potential is then obtained by appropriate substitutions into equation(11).

APPENDIX C

BACKWARD TIME STOCHASTIC INTEGRATOR

The simulations of the reaction dynamics presented in Section 5.5 require one to follow a stochastic trajectory numerically from $t = 0$ both forward in time to $t = T/2$ and backward in time to $t = -T/2$. For the forward propagation, a standard stochastic integration scheme [139, 140] has been implemented. The backward integration requires special care if one wishes to follow the same stochastic trajectory both forward and backward in time. The modification of the integration scheme that is necessary to this end is described here.

The forward numerical integrator for Langevin equations [139, 140] takes the form

$$\begin{aligned} r(t+\delta t) &= r(t) + c_1 v(t) + c_2 a(t) + \delta r \ , && (87) \\ v(t+\delta t) &= c_3 v(t) + c_4 a(t) + c_5 a(t+\delta t) + \delta v \ , && (88) \end{aligned}$$

where $a(t)$ is the acceleration caused by the potential of mean force, the c_i are numerical coefficients that depend on the time step δt and the damping constant γ, and the random variables δr and δv are sampled from a known Gaussian distribution. Time reversal in this algorithm can be obtained through a shift in time by $-\delta t$ so that t becomes $t - \delta t$ and $t + \delta t$ becomes t. This replacement and a simple reorganization leads to

$$\begin{aligned} r(t-\delta t) &= r(t) - c_1 v(t-\delta t) - c_2 a(t-\delta t) - \delta r \ , && (89) \\ v(t-\delta t) &= \frac{1}{c_3}\left[v(t) - c_4 a(t-\delta t) - c_5 a(t) - \delta v\right] \ . && (90) \end{aligned}$$

The backward step (89) cannot be evaluated as it stands because the acceleration $a(t - \delta t)$ depends on the position $r(t - \delta t)$ that is yet to be determined. To circumvent this problem, we substitute equation (90) into equation (89) to obtain

$$r(t-\delta t) = r(t) - \frac{c_1}{c_3}\left[v(t) - c_5 a(t) - \delta v\right] - \delta r + \left(\frac{c_1 c_4}{c_3} - c_2\right) a(t-\delta t) \ . \qquad (91)$$

When the acceleration $a(t - \delta t)$ is expressed in terms of the position $r(t - \delta t)$ through the equation of motion, equation (91) becomes an implicit equation for the positions $r(t - \delta t)$

at the earlier time. For all but the simplest potentials, it cannot be solved explicitly. Specifically, for the anharmonic potential (34),

$$U(x,y) = -\frac{1}{2}\omega_x^2 x^2 + \frac{1}{2}\omega_y^2 y^2 + kx^2y^2 \ , \tag{92}$$

it leads to the coupled equation system

$$\begin{aligned} x(t-\delta t) &= X(t) + \left(\frac{c_1c_4}{c_3} - c_2\right)\left(\omega_x^2 x(t-\delta t) - 2kx(t-\delta t)y(t-\delta t)^2\right) \ , &(93)\\ y(t-\delta t) &= Y(t) - \left(\frac{c_1c_4}{c_3} - c_2\right)\left(\omega_y^2 y(t-\delta t) + 2kx(t-\delta t)^2y(t-\delta t)\right) \ , &(94)\end{aligned}$$

where $X(t)$ and $Y(t)$ denote the contributions of the first three terms in equation (91). In the harmonic limit $k = 0$, the two equations uncouple and can be solved for simple explicit expressions for the position updates. For nonzero k, equations (93) and (94) represent an implicit integration scheme. It can be converted into an explicit method by rearranging the terms into

$$\begin{aligned} x(t-\delta t) &= \frac{X(t)}{1-\left(\frac{c_1c_4}{c_3}-c_2\right)\left(\omega_x^2 - 2ky(t-\delta t)^2\right)} \ , &(95)\\ y(t-\delta t) &= \frac{Y(t)}{1+\left(\frac{c_1c_4}{c_3}-c_2\right)\left(\omega_y^2 + 2kx(t-\delta t)^2\right)} \ . &(96)\end{aligned}$$

The denominators in equations (95) and (96) are updated using equations (93) and (94), but the unknown corrections involving $a(t-\delta t)$ are neglected because the coefficient $c_1c_4/c_3 - c_2$ is of second order in the time step δt. This leads to

$$\begin{aligned} x(t-\delta t) &\approx \frac{X(t)}{1-\left(\frac{c_1c_4}{c_3}-c_2\right)\left(\omega_x^2 - 2kY(t)^2\right)} \ , &(97)\\ y(t-\delta t) &\approx \frac{Y(t)}{1+\left(\frac{c_1c_4}{c_3}-c_2\right)\left(\omega_y^2 + 2kX(t)^2\right)} \ . &(98)\end{aligned}$$

Finally, we insert these approximations into the right-hand sides of equations (95) and (96) to obtain an explicit integration scheme backwards in time.

APPENDIX D

CONSTRUCTION OF THE DIHEDRAL-ANGLE DISTRIBUTIONS

This appendix describes the specific approach used to generate the ψ_i-ϕ_{i+1} and ϕ_i-ψ_i plots. An analysis of these distributions, confirming that the former is not trivially related to the latter, is also included here. Additionally the construction of the uniformizing cumulative Gaussian distribution used to generate the D_2 check is provided.

D.0.3 Data-mining the ψ_i-ϕ_{i+1} distributions

All 400 ψ_i-ϕ_{i+1} plots have been generated for each of the five sublibraries, NR50, NR70, NR90, NR100 and EXP. The ψ_i-ϕ_{i+1} space has been discretized into a two-dimensional grid consisting of $5^\circ \times 5^\circ$ bins. Often, such results are summarized in a plot of the frequencies of occurrence, that is, the number of observed angles in a given bin [248, 234]. If the sample size is large enough, such a frequency (normalized by all the hits) is a good estimate of the integrated probability over all structures with angles in the domain space of the bin. In what follows, the sampling size of the sublibraries of the current PDB is assumed to be "large enough" and this integrated probability is loosely referred to as a probability. In addition, the 20 ϕ_i-ψ_i plots are generated from the pruned sublibrary following the same approach. These 420 histogramed distributions can be saved into a single database which can, in turn, be used to calculate the structure entropy differences as described in the text.

The dependence of ψ_i and ϕ_{i+1} on the identity of the residue can be further elucidated by comparison with the one-dimensional distributions obtained from projections of the

Ramachandran and ψ_i-ϕ_{i+1} plots, *i.e.*,

$$\begin{aligned} P'_R(\psi_i) &= \int d\phi_i P_R(\phi_i, \psi_i) && \text{(99a)} \\ P''_R(\phi_i) &= \int d\psi_i P_R(\phi_i, \psi_i) && \text{(99b)} \\ P'_{R_<, R_>}(\psi_i) &= \int d\phi_{i+1} P_{R_<, R_>}(\psi_i, \phi_{i+1}) && \text{(99c)} \\ P''_{R_<, R_>}(\phi_{i+1}) &= \int d\psi_i P_{R_<, R_>}(\psi_i, \phi_{i+1}) \, . && \text{(99d)} \end{aligned}$$

In addition, a measure of the correlation between ψ_i and ϕ_{i+1} can be obtained from the difference,

$$P_c = |P_{R_<, R_>}(\psi_i, \phi_{i+1}) - P'_{R_<}(\psi_i) P'_{R_>}(\phi_{i+1})| \, . \tag{100}$$

A nonzero value of P_c indicates correlation between the two angles.

D.0.4 Dihedral Angle Distributions

The $\psi_i - \phi_i$ and $\phi_i - \psi_{i+1}$ dihedral angle distibutions have been obtained for all five libraries described in Section D.0.3. The observations to be described in this subsection are true for each of these sets and hence will be illustrated using only one of them; namely the NR90 sublibrary. Although any of the others could be used for this purpose, it has also been chosen because it was found to be the best choice (as described in Section 6.3.2 of the primary manuscript) for the use of the D_1 and D_2 scores defined in Section 6.3.3, and hence this illustrations is also useful in interpreting the subsequent analysis of the manuscript.

The relative abundance of each residue pair and that of each residue in the NR90 sublibrary are shown in Table 7. Eight representative plots of ψ_i-ϕ_{i+1} distributions are shown in Figure 38. All 20 ϕ_i-ψ_i and 400 ψ_i-ϕ_{i+1} plots are included in the Supplement B. The chosen distributions are those for the $\{i, (i+1)\}$ pairs, ALA-SER *(a)*, ALA-GLU *(b)*, SER-ALA *(c)*, GLU-ALA *(d)*, ARG-GLY *(e)*, ARG-GLU *(f)*, GLY-ARG *(g)*, and GLU-ARG *(h)*. It is easily seen that the distribution varies in a nontrivial manner for each residue. In particular, each row illustrates that the distribution is heavily dependent on the order of the sequence. (This is not surprising since the structure along the dipeptide bond is strongly influenced by the order of the residues.) In addition, the differences between the two pairs of distributions —*a,* b and *e, f*— in which the second residue is varied while the other remains the constant, as well as the two pairs of distributions —*c, d* and *g, h*— for which the first residue is similarly varied, also demonstrate that the identity of a distant residue affects the ψ_i-ϕ_{i+1} distributions.

Figure 39 shows the correlation, P_c, calculated from equation (100) between ψ_i and ϕ_{i+1} for two pairs ALA-GLU and ARG-GLU. It illustrates that there is strong correlation between the two angles, which provides further justification that the ψ_i-ϕ_{i+1} plot contains nontrivial information. Thus both the ϕ_i-ψ_i and ψ_i-ϕ_{i+1} plots provide important information about the conformation of protein secondary structures.

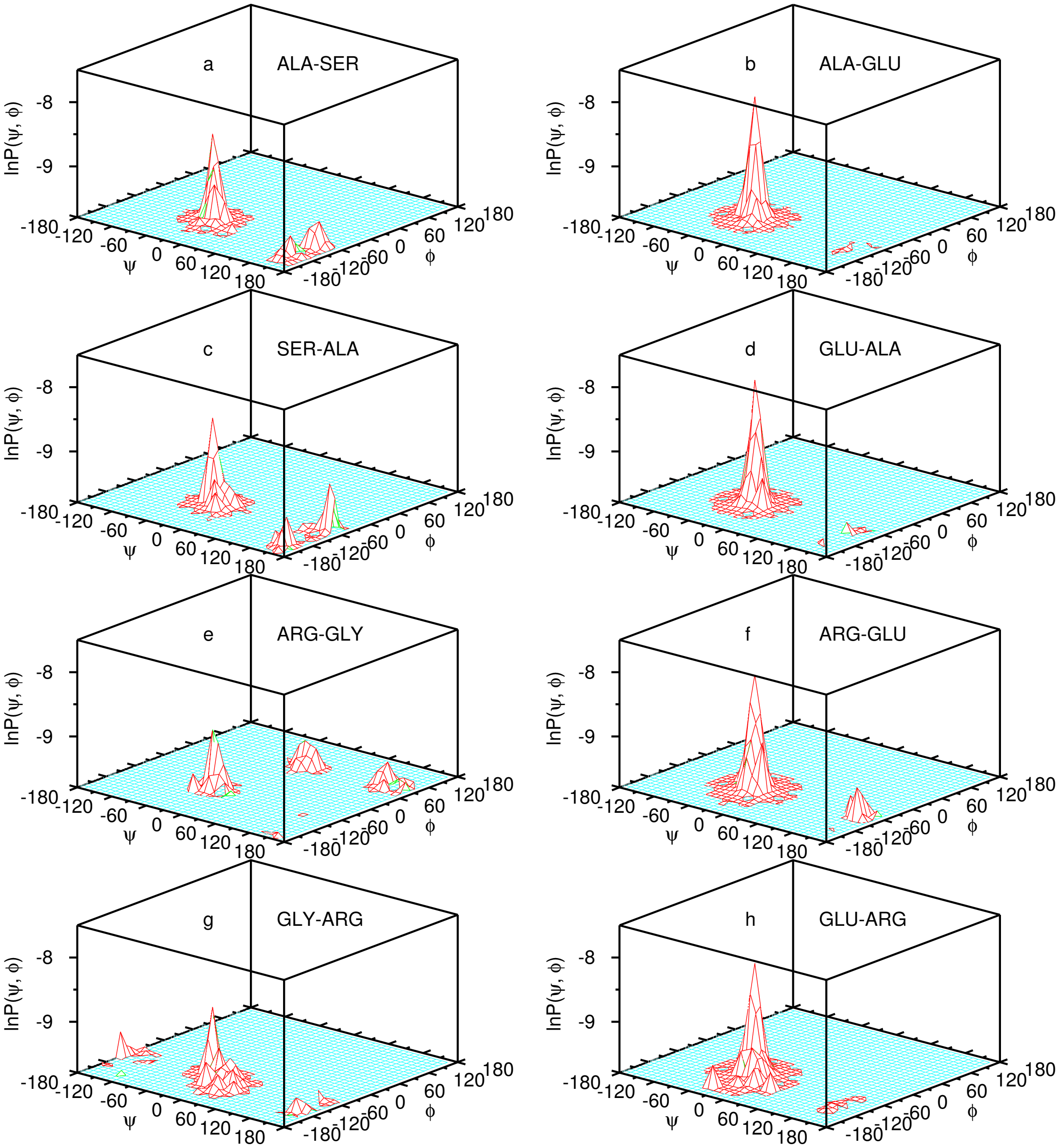

Figure 38: Logarithmic probability distribution of ψ_i-ϕ_{i+1} for ALA-SER (a); ALA-GLU (b); SER-ALA (c); GLU-ALA (d); ARG-GLY (e); ARG-GLU (f); GLY-ARG (g); GLU-ARG (h). The results were obtained from samples of 2,696; 2,517; 2,424; 2,532; 1,520; 1,494; 1,844; 1,501 angle pairs, respectively. Three-letter names are used to identify the corresponding amino acid as in Table 7. The subscripts i and $i+1$ on ψ_i and ϕ_{i+1}, respectively, were omitted. The scale is the same for all plots.

Table 7: The relative abundance that a given ordered $R_<$-$R_>$ residue pair is found in the NR90 sublibrary for all 400 possible such pairs is listed below. Note that the ordering is from the N' to C', and hence the abundance of $R_<$-$R_>$ is not necessarily equal to that of $R_>$-$R_<$. Three-letter names are used to identify the corresponding amino acid. The total number of pairs in the database is 506,633. The last line displays the relative abundance of given residue in the database for all 20 amino acids. The total number of residues in the database is 503,871.

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	0.76	0.36	0.32	0.45	0.12	0.29	0.50	0.59	0.18	0.41	0.67	0.46	0.16	0.28	0.33	0.53	0.45	0.11	0.26	0.56
Arg	0.33	0.25	0.19	0.27	0.09	0.17	0.30	0.30	0.10	0.28	0.41	0.23	0.09	0.23	0.20	0.25	0.21	0.06	0.18	0.35
Asn	0.32	0.18	0.25	0.22	0.08	0.16	0.24	0.40	0.10	0.27	0.38	0.23	0.08	0.20	0.26	0.31	0.27	0.08	0.19	0.31
Asp	0.45	0.24	0.25	0.30	0.11	0.17	0.35	0.47	0.11	0.35	0.50	0.35	0.11	0.27	0.29	0.35	0.33	0.09	0.26	0.40
Cys	0.15	0.10	0.10	0.10	0.07	0.08	0.10	0.15	0.05	0.08	0.16	0.12	0.03	0.07	0.11	0.14	0.11	0.02	0.06	0.10
Gln	0.30	0.17	0.16	0.18	0.07	0.20	0.22	0.29	0.07	0.19	0.34	0.22	0.08	0.15	0.17	0.26	0.21	0.07	0.14	0.27
Glu	0.50	0.30	0.29	0.36	0.10	0.24	0.48	0.41	0.12	0.38	0.57	0.43	0.13	0.24	0.22	0.31	0.31	0.12	0.22	0.44
Gly	0.57	0.36	0.34	0.45	0.16	0.29	0.41	0.64	0.16	0.43	0.58	0.50	0.14	0.32	0.30	0.62	0.52	0.12	0.33	0.54
His	0.14	0.09	0.09	0.11	0.06	0.08	0.11	0.17	0.09	0.12	0.18	0.11	0.05	0.10	0.14	0.13	0.13	0.05	0.08	0.14
Ile	0.44	0.25	0.28	0.34	0.10	0.19	0.34	0.37	0.12	0.30	0.38	0.36	0.09	0.19	0.26	0.39	0.35	0.07	0.20	0.36
Leu	0.67	0.44	0.37	0.47	0.13	0.39	0.58	0.60	0.17	0.39	0.65	0.56	0.14	0.27	0.39	0.60	0.53	0.08	0.27	0.57
Lys	0.49	0.24	0.29	0.37	0.11	0.21	0.43	0.43	0.11	0.36	0.54	0.42	0.11	0.23	0.27	0.33	0.34	0.08	0.22	0.46
Met	0.15	0.09	0.09	0.12	0.03	0.08	0.13	0.15	0.04	0.10	0.16	0.15	0.07	0.07	0.07	0.13	0.12	0.02	0.06	0.14
Phe	0.27	0.16	0.19	0.27	0.07	0.15	0.22	0.30	0.09	0.19	0.31	0.25	0.07	0.14	0.24	0.28	0.27	0.05	0.16	0.24
Pro	0.38	0.20	0.20	0.28	0.08	0.16	0.37	0.42	0.10	0.20	0.35	0.24	0.07	0.17	0.22	0.39	0.24	0.07	0.16	0.36
Ser	0.48	0.28	0.29	0.36	0.16	0.25	0.34	0.74	0.13	0.34	0.57	0.35	0.12	0.27	0.29	0.73	0.40	0.12	0.25	0.49
Thr	0.46	0.23	0.25	0.32	0.14	0.20	0.30	0.46	0.13	0.34	0.58	0.31	0.10	0.28	0.30	0.40	0.39	0.12	0.25	0.53
Trp	0.10	0.06	0.09	0.09	0.03	0.07	0.09	0.14	0.04	0.09	0.11	0.10	0.03	0.05	0.06	0.09	0.09	0.03	0.06	0.12
Tyr	0.25	0.17	0.20	0.23	0.10	0.15	0.20	0.28	0.08	0.19	0.29	0.18	0.07	0.19	0.20	0.26	0.25	0.09	0.20	0.22
Val	0.56	0.33	0.32	0.45	0.13	0.25	0.44	0.44	0.17	0.37	0.58	0.49	0.13	0.26	0.33	0.46	0.56	0.09	0.25	0.52
	7.78	4.49	4.56	5.73	1.90	3.76	6.14	7.76	2.17	5.39	8.28	6.05	1.88	3.94	4.64	6.94	6.11	1.55	3.79	7.12

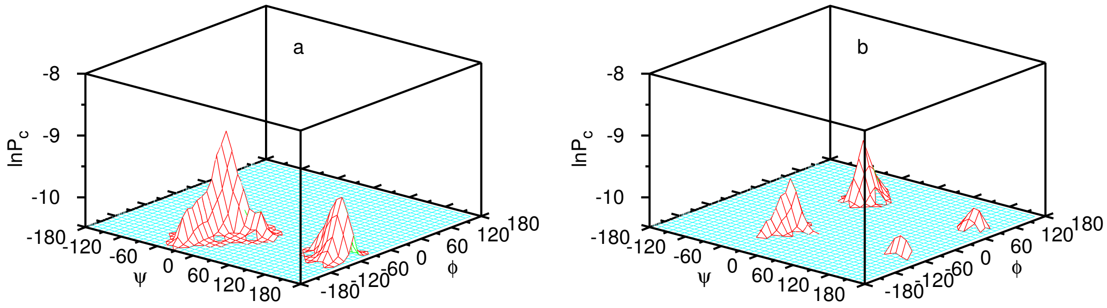

Figure 39: Correlation between ψ_i and ϕ_{i+1} for ALA-GLU (a) and ARG-GLY (b). P_c is on a logarithmic scale and the subscripts on ψ_i and ϕ_{i+1} were omitted for clarity.

D.0.5 The D_2 checking function for a Gaussian Distribution

If the probability distribution in equation (71) were a Gaussian distribution with median located at zero, and standard deviation σ, then both it and D_2 check in equation (73) could be measured directly with only a slight modification of the domain of integration. Namely,

$$I(x) \equiv \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2\sigma^2} dy & \text{if } x < 0 \\ \frac{1}{\sigma\sqrt{2\pi}} \int_{x}^{\infty} e^{-y^2/2\sigma^2} dy & \text{if } x \geq 0 \end{cases} , \tag{101}$$

which can also be expressed by the error function,

$$I(x) \equiv \begin{cases} \frac{1}{2}(1 + \mathrm{erf}(\frac{x}{\sigma\sqrt{2}})) & \text{if } x < 0 \\ \frac{1}{2}(1 - \mathrm{erf}(\frac{x}{\sigma\sqrt{2}})) & \text{if } x \geq 0 \end{cases} , \tag{102}$$

where erf(·) is the error function defined as

$$\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \; . \tag{103}$$

If I is expressed in terms of the number of standard deviations away from the median as

$$I(\pm n\sigma) \equiv \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^{n\sigma} e^{-y^2/2\sigma^2} dy \; , \tag{104}$$

then insertion of I into equation (73) leads to the claimed result,

$$D_2 = \pm n\sigma \; . \tag{105}$$

By setting σ equal to 1 in the definition of D_2 as done in equation (73), D_2 takes on values that are equal to the increment in standard deviations for the same relative cumulative probability as in the Gaussian distribution. This mapping performed by way of equating cumulative probabilities effectively "uniformizes" the actual distribution.

REFERENCES

[1] Gomer, R., Diffusion of adsorbates on metal surfaces, *Rep. Prog. Phys.*, vol. 53, pp. 917–1002, 1990.

[2] Kellogg, G. L., Field-ion microscope studies of single-atom surface-diffusion and cluster nucleation on metal surfaces, *Surf. Sci. Rep.*, vol. 21, p. 1, 1994.

[3] Barth, J. V., Transport of adsorbates at metal surfaces: From thermal migration to hot precursors, *Surf. Sci. Rep.*, vol. 40, pp. 75–149, 2000.

[4] Naumovets, A. G. and Zhang, Z., Fidgety particles on surfaces: how do they jump, walk, group, and settle in virgin areas?, *Surf. Sci.*, vol. 500, pp. 414–436, 2002.

[5] Naumovets, A. G., Collective surface diffusion: An experimentalist's view, *Physica A*, vol. 357, pp. 189–215, 2005.

[6] Graham, A. P., The low energy dynamics of adsorbates on metal surfaces investigated with helium atom scattering, *Surf. Sci. Rep*, vol. 49, pp. 115–168, 2003.

[7] Tully, J. C., Theories of the dynamics of inelastic and reactive processes at surfaces, *Annu. Rev. Phys. Chem.*, vol. 31, pp. 319–343, 1980.

[8] Doll, J. and Voter, A. F., Recent development in the theory of surface diffusion, *Annu. Rev. Phys. Chem.*, vol. 38, pp. 413–431, 1987.

[9] Feibelman, P. J., Theory of adsorbate interactions, *Annu. Rev. Phys. Chem.*, vol. 40, pp. 261–290, 1989.

[10] Tully, J. C., Chemical dynamics at metal surfaces, *Annu. Rev. Phys. Chem.*, vol. 51, pp. 153–178, 2000.

[11] Ala-Nissila, T., Ferrando, R., and Ying, S. C., Collective and single particle diffusion on surfaces, *Adv. Phys.*, vol. 51, p. 949, 2002.

[12] Miret-Artés, S. and Pollak, E., The dynamics of activated surface diffusion, *J. Phys.: Condens. Matter*, vol. 17, pp. S4133–S4135, 2005.

[13] Starrost, F. and Carter, E. A., Modeling the full monty: baring the nature of surfaces across time and space, *Surf. Sci.*, vol. 500, p. 323, 2002.

[14] Voter, A. F., Montalenti, F., and Germann, T. C., Extending the time scale in atomistic simulations of materials, *Annu. Rev. Mater. Res.*, vol. 32, pp. 321–346, 2002.

[15] Haight, R., Electron dynamics at surfaces, *Surf. Sci. Rep*, vol. 21, pp. 275–325, 1995.

[16] Radeke, M. R. and Carter, E. A., *Ab initio* dynamics of surface chemistry, *Annu. Rev. Phys. Chem.*, vol. 53, pp. 248–270, 1997.

[17] Brivio, G. P. and Trioni, M. I., The adiabatic molecule-metal surface interaction: Theoretical approaches, *Rev. Mod. Phys.*, vol. 71, p. 231, 1999.

[18] Saalfrank, P., Quantum dynamical approach to ultrafast molecular desorption from surfaces, *Chem. Rev.*, vol. 106, pp. 4116–4159, 2006.

[19] Tully, J. C., Mixed quantum-classical dynamics, *Faraday Disc. Chem. Soc.*, vol. 110, pp. 407–419, 1998.

[20] Head-Gordon, M. and Tully, J. C., Molecular dynamics with electronic frictions, *J. Chem. Phys.*, vol. 103, p. 10137, 1995.

[21] Kindt, J. T., Tully, J. C., Head-Gordon, M., and Gomez, M. A., Electron-hole pair contributions to scattering, sticking, and surface diffusion: CO on Cu(100), *J. Chem. Phys.*, vol. 109, p. 3629, 1998.

[22] Wodtke, A. M., Tully, J. C., and Auerbach, D. J., Electronically non-adiabatic interactions of molecules at metal surfaces: Can we trust the Born-Oppenheimer approximation for surface chemistry?, *Int. Rev. Phys. Chem.*, vol. 23, pp. 513–539, 2004.

[23] Hofmann, F. and Toennies, J. P., High-resolution helium atom time-of-flight spectroscopy of low-frequency vibrations of adsorbates, *Chem. Rev.*, vol. 96, pp. 1307–1326, 1996.

[24] Swartzentruber, B. S., Direct measurement of surface diffusion using atom-tracking scanning tunneling microscopy, *Phys. Rev. Lett.*, vol. 76, p. 459, 1996.

[25] Zambelli, T., Trost, J., Wintterlin, J., and Ertl, G., Diffusion and atomic hopping of N atoms on Ru(0001) studied by scanning tunneling microscopy, *Phys. Rev. Lett.*, vol. 76, p. 795, 1996.

[26] Pedersen, M. Ø., Österlund, L., Mortensen, J. J., Mavrikakis, M., Hansen, L. B., Stensgaard, I., Lægsgaard, E., Nørskov, J. K., and Besenbacher, F., Diffusion of N Adatoms on the Fe(100) Surface, *Phys. Rev. Lett.*, vol. 84, p. 4898, 2000.

[27] Jardine, A., Ellis, J., and Allison, W., Quasi-elastic helium-atom scattering from surfaces: experiment and interpretation, *J.Phys.:Condens.Matter*, vol. 14, p. 6173, 2002.

[28] Graham, A. P., Hofmann, F., Toennies, J. P., Chen, L. Y., and Ying, S. C., Experimental and theoretical investigation of the microscopic vibrational and diffusional dynamics of sodium atoms on a Cu(001) surface, *Phys. Rev. B*, vol. 56, p. 10567, 1997.

[29] Ellis, J., Graham, A. P., Hofmann, F., and Toennies, J. P., Coverage dependence of the microscopic diffusion of Na atoms on the Cu(001) surface: A combined helium atom scattering experiment and molecular dynamics study, *Phys. Rev. B*, vol. 63, p. 195408, 2001.

[30] Agrawal, P. M., Rice, B. M., and Thompson, D. L., Predicting trends in rate parameters for self-diffusion on FCC metal surfaces, *Surf. Sci.*, vol. 515, pp. 21–35, 2002.

[31] Feibelman, P. J., Scaling of hopping self-diffusion barriers on fcc (100) surfaces with bulk bond energies, *Surf. Sci.*, vol. 423, p. 169, 1999.

[32] Kürprik, U., Kara, A., and Rahman, T. S., Role of lattice vibrations in adatom diffusion, *Phys. Rev. Lett.*, vol. 78, p. 1086, 1997.

[33] Tully, J. C., Gilmer, G. H., and Shugard, M., Molecular dynamics of surface diffusion. I. The motion of adatoms and clusters, *J. Chem. Phys.*, vol. 71, p. 1630, 1979.

[34] Tully, J. C., Dynamics of gas-surface interactions: 3D generalized Langevin model applied to fcc and bcc surfaces, *J. Chem. Phys.*, vol. 73, p. 1975, 1980.

[35] Tsekov, R. and Ruckenstein, E., Stochastic dynamics of a subsystem interacting with a solid body with application to diffusive processes in solids, *J. Chem. Phys.*, vol. 100, p. 1450, 1994.

[36] Tsekov, R. and Ruckenstein, E., Two-dimensional Brownian motion of atoms and dimers on solid surfaces, *Surf. Sci.*, vol. 344, p. 175, 1995.

[37] Georgievski, Y., Kozhushner, M. A., and Pollak, E., Activated surface diffusion: Are correlated hops the rule or the exception?, *J. Chem. Phys.*, vol. 102, pp. 6908–6918, 1995.

[38] Gershinsky, G., Georgievskii, Y., Pollak, E., and Betz, G., Microscopic and macroscopic estimates of friction: application to surface diffusion of copper, *Surf. Sci.*, vol. 365, pp. 159–167, 1996.

[39] Pollak, E., Bader, J., Berne, B. J., and Talkner, P., Theory of correlated hops in surface diffusion, *Phys. Rev. Lett.*, vol. 70, p. 3299, 1993.

[40] Georgievski, Y. and Pollak, E., Long hops of an adatom on a surface, *Surf. Sci. Lett.*, vol. 355, pp. L366–L370, 1996.

[41] Wahnström, G., Role of phonons and electron-hole pairs in hydrogen diffusion on a corrugated metal surface, *Chem. Phys. Lett.*, vol. 163, p. 401, 1989.

[42] Sholl, D. S. and Skodje, R. T., Kinetic phase transitions and bistability in a model surface reaction II. Spatially inhomogeneous theories, *Surf. Sci.*, vol. 334, p. 305, 1995.

[43] Brown, D. E., Sholl, D. S., Skodje, R. T., and George, S. M., Surface diffusion of H and CO on Cu/Ru(001): evidence for long-range trapping by copper islands, *Chem. Phys.*, vol. 205, p. 23, 1996.

[44] Chandler, D., Statistical mechanics of isomerization dynamics in liquids and the transition state approximation, *J. Chem. Phys.*, vol. 68, pp. 2959–2970, 1978.

[45] Voter, A. F. and Doll, J. D., Dynamical correction to transition state theory for multistate systems, *J. Chem. Phys.*, vol. 82, pp. 80–92, 1985.

[46] Zhang, Z. and Metiu, H., Adsorbate migration on a solid surface: The connection between hopping dynamics and the atom-surface interaction energy, *J. Chem. Phys.*, vol. 93, p. 2087, 1990.

[47] Hänggi, P., Talkner, P., and Borkovec, M., Reaction-Rate Theory: Fifty Years After Kramers, *Rev. Mod. Phys.*, vol. 62, pp. 251–341, 1990. and references therein.

[48] Pollak, E. and Talkner, P., Reaction rate theory: What it was, where it is today, and where is it going?, *Chaos*, vol. 15, pp. 026116–1–11, 2005.

[49] Ferrón, J., Gómez, L., de Miguel, J. J., and Miranda, R., Nonstochastic behavior of atomic surface diffusion on Cu(111) down to low temperatures, *Phys. Rev. Lett.*, vol. 93, p. 166107, 2004.

[50] Kellogg, G. L. and Feibelman, P. J., Surface self-diffusion on Pt(001) by atomic exchange mechansim, *Phys. Rev. Lett.*, vol. 64, p. 3143, 1990.

[51] Feibelman, P. J., Diffusion path for an Al adatom on Al(001), *Phys. Rev. Lett.*, vol. 65, pp. 729–732, 1990.

[52] Adelman, S. A. and Doll, J. D., Generalized Langevin equation approach for atom/solid-surface scattering: Collinear atom/harmonic chain model, *J. Chem. Phys.*, vol. 61, pp. 4242–45, 1974.

[53] Nitzan, A., Shugard, M., and Tully, J. C., Stochastic classical trajectory approach to relaxation phenomena. II. Vibrational relaxation of impurity molecules in Debye solids, *J. Chem. Phys.*, vol. 69, p. 2525, 1978.

[54] Tully, J. C. and Cardillo, M. J., Dynamics of molecular motion at single-crystal surfaces, *Science*, vol. 223, pp. 445–450, 1984.

[55] Rettner, C. T., Auerbach, D. J., Tully, J. C., and Kleyn, A. W., Chemical dynamics at the gas-surface interface, *J. Phys. Chem.*, vol. 100, pp. 13021–13033, 1996.

[56] Tully, J. C., Washboard model of gas-surface scattering, *J. Chem. Phys.*, vol. 92, p. 680, 1990.

[57] Yan, T., Hase, W. L., and Tully, J. C., A washboard with moment of inertia model of gas-surface scattering, *J. Chem. Phys.*, vol. 120, p. 1031, 2004.

[58] Tully, J. C., Gomez, M., and Head-Gordon, M., Electronic and phonon mechanisms of vibrational relaxation: CO on Cu(100), *J. Vac. Sci. Technol.*, vol. 11, p. 1914, 1993.

[59] Yildirim, H., Kara, A., Durukanoğlu, S., and Rahman, T. S., Calculated pre-exponential factors and energetics for adatom hopping on terraces and steps of Cu(100) and Cu(110), *Surf. Sci.*, vol. 600, pp. 484–492, 2006.

[60] Durukanoğlu, S., Kara, A., and Rahman, T. S., The role of lattice vibrations in adatom diffusion at metal stepped surfaces, *Surf. Sci.*, vol. 587, p. 128, 2005.

[61] Al-Rawi, A. H., Kara, A., and Rahman, T. S., Comparative study of anharmonicity: Ni(111), Cu(111), and Ag(111), *Phys. Rev. B*, vol. 66, p. 165439, 2002.

[62] Haug, K., Wahnström, G., and Metiu, H., Hydrogen motion on a Cu surface: A model study of the rate of single and double site-to-site jumps and the role of the motion perpendicular to the surface, *J. Chem. Phys.*, vol. 90, p. 540, 1989.

[63] Haug, K., Wahnström, G., and Metiu, H., Hydrogen motion on a rigid Cu surface: The calculation of the site to site hopping rate by using flux-flux correlation functions, *J. Chem. Phys.*, vol. 92, p. 2083, 1990.

[64] Haug, K. and Metiu, H., Quantum simulation of hydrogen migration on Ni(100): The role of fluctuations, recrossing, and multiple jumps, *J. Chem. Phys.*, vol. 94, p. 3251, 1991.

[65] Rick, S. W., Lynch, D. L., and Doll, J. D., The quantum dynamics of hydrogen and deuterium on the Pd(lll) surface: A path integral transition state theory study, *J. Chem. Phys.*, vol. 99, p. 8183, 1993.

[66] Carter, E. A., Adsorption and diffusion energetics of hydrogen atoms on Fe(110) from first principles, *Surf. Sci.*, vol. 547, p. 85, 2003.

[67] Evangelakis, G. A. and Papanicolaou, N. I., Adatom self-diffusion processes on (001) copper surface by molecular dynamics, *Surf. Sci.*, vol. 347, pp. 376–386, 1996.

[68] Kallinteris, G. C., Evangelakis, G. A., and Papanicolaou, N. I., Molecular dynamics study of the vibrational and transport properties of copper adatoms on the (111) copper surface; comparison with the (001) face, *Surf. Sci.*, vol. 369, pp. 185–198, 1996.

[69] Papanicolaou, N. I., Papathanakos, V. C., and Papageorgiou, D. G., Self-diffusion on Al(100) and Al(111) surfaces by molecular-dynamics simulation, *Physica B*, vol. 296, pp. 259–263, 2001.

[70] Prévot, G., Cohen, C., Schmaus, D., and Pontikis, V., Non-isotropic surface diffusion of lead on Cu(110): a molecular dynamics study, *Surf. Sci.*, vol. 459, pp. 57–68, 2000.

[71] Bulou, H., Lucas, O., Kibaly, M., and Goyhenex, C., Long-time scale molecular dynamics study of Co diffusion on the Au(111) surface, *Comp. Mat. Sci*, vol. 27, pp. 181–185, 2003.

[72] Ramirez-Pastor, A. J., Nazzarro, M., Riccardo, J. L., and Pereya, V., Surface diffusion of dimers: I repulsive interactions, *Surf. Sci.*, vol. 391, pp. 267–277, 1997.

[73] Papathanakos, V. and Evangelakis, G. A., Structural and diffusive properties of small 2D Au clusters on the Cu(111) surface, *Surf. Sci.*, vol. 499, pp. 229–243, 2002.

[74] Jacobsen, J., Jacobsen, K. W., and Sethna, J. P., Rate theory for correlated processes: Double jumps in adatom diffusion, *Phys. Rev. Lett.*, vol. 79, p. 2843, 1997.

[75] Sørensen, M. R. and Voter, A. F., Temperature-accelerated dynamics for simulation of infrequent events, *J. Chem. Phys.*, vol. 112, p. 9599, 2000.

[76] Miron, R. A. and Fichthorn, K. A., Accelerated molecular dynamics with the bond-boost method, *J. Chem. Phys.*, vol. 119, p. 6210, 2003.

[77] Daw, M. S. and Baskes, M. I., Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals, *Phys. Rev. Lett.*, vol. 50, p. 1285, 1983.

[78] Daw, M. S. and Baskes, M. I., Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals, *Phys. Rev. B*, vol. 29, p. 6443, 1984.

[79] Jacobsen, K. W., Nørskov, J. K., and Puska, M. J., Interatomic interactions in the effective-medium theory, *Phys. Rev. B*, vol. 35, p. 7423, 1987.

[80] Bae, C., Freeman, D. L., Doll, J. D., Kresse, G., and Hafner, J., Energetics of hydrogen chemisorbed on Cu(110): A first principles calculations study, *J. Chem. Phys.*, vol. 113, p. 6926, 2000.

[81] Mori, H., Transport, collective motion, and brownian motion, *Prog. Theor. Phys.*, vol. 33, pp. 423–455, 1965.

[82] Zwanzig, R., Ensemble method in the theory of irreversibility, *Phys. Rev.*, vol. 124, p. 983, 1961.

[83] Zwanzig, R., Statistical mechanics of irreversibility, in *Lectures in Theoretical Physics (Boulder)* (Brittin, W. E., Downs, B. W., and Downs, J., eds.), vol. 3, pp. 106–141, New York: Wiley-Interscience, 1961.

[84] Zwanzig, R., *Nonequilibrium Statistical Mechanics.* London: Oxford University Press, 2001.

[85] Kubo, R., The fluctuation-dissipation theorem, *Rep. Prog. Phys.*, vol. 29, pp. 255–284, 1966.

[86] Moix, J. and Hernandez, R. unpublished results.

[87] Bao, J. D., Abe, Y., and Zhuo, Y. Z., Inhomogeneous friction leading to current in periodic system, *Physica A*, vol. 265, pp. 111–118, 1999.

[88] Zhang, X. P. and Bao, J. D., Stochastic resonance in multidimensional periodic potential, *Surf. Sci.*, vol. 540, p. 145, 2003.

[89] Talkner, P., Hershkovitz, E., Pollak, E., and Hänggi, P., Controlling activated surface diffusion by external fields, *Surf. Sci.*, vol. 437, pp. 198–206, 1999.

[90] Moix, J. M., Shepherd, T. D., and Hernandez, R., A phenomenological model for surface diffusion: Diffusive dynamics across incoherent stochastic aperiodic potentials, *J. Phys. Chem. B*, vol. 108, pp. 19476–19482, 2004.

[91] Chen, L. Y., Baldan, M. R., and Ying, S. C., Surface diffusion in the low-friction limit: Occurence of long jumps, *Phys. Rev. B*, vol. 54, p. 8856, 1996.

[92] Cucchetti, A. and Ying, S. C., Memory effects in the frictional damping of diffusive and vibrational motion of adatoms, *Phys. Rev. B*, vol. 54, p. 3300, 1996.

[93] Moro, G. J. and Polimeno, A., Multi-barrier crossing regulated by the friction, *Chem. Phys. Lett.*, vol. 189, pp. 133–137, 1992.

[94] Vega, J. L., Guantes, R., and Miret-Artés, S., Mean first passage time and the Kramers turnover theory in activated atom-surface diffusion, *Phys. Chem. Chem. Phys.*, vol. 4, p. 4985, 2002.

[95] Sancho, J. M., Lacasta, A. M., Lindenberg, K., Sokolov, I. M., and Romero, A. H., Diffusion on a solid surface: Anomalous is normal, *Phys. Rev. Lett.*, vol. 92, p. 250601, 2004.

[96] Lacasta, A. M., Sancho, J. M., Romero, A. H., Sokolov, I. M., and Lindenberg, K., From subdiffusion to superdiffusion of particles on solid surfaces, *Phys. Rev. E*, vol. 70, p. 051104, 2004.

[97] Hershkovitz, E., Talkner, P., Pollak, E., and Georgievski, Y., Multiple hops in multidimensional activated surface diffusion, *Surf. Sci.*, vol. 421, pp. 73–88, 1999.

[98] Montalenti, F. and Ferrando, R., Jumps and concerted moves in Cu, Ag, and Au(110) adatom self-diffusion, *Phys. Rev. B*, vol. 59, p. 5881, 1999.

[99] Braun, O. M. and Ferrando, R., Role of long jumps in surface diffusion, *Phys. Rev. E*, vol. 65, p. 061107, 2002.

[100] Ferrando, R., Spadacini, R., and Tommei, G. E., Kramers problem in periodic potentials: Jump rate and jump lengths, *Phys. Rev. E*, vol. 48, p. 2437, 1993.

[101] Mattsson, T. R. and Wahnström, G., Isotope effect in hydrogen surface diffusion, *Phys. Rev. B*, vol. 56, p. 14944, 1997.

[102] Hernandez, R., The projection of a mechanical system onto the irreversible generalized Langevin equation (iGLE), *J. Chem. Phys.*, vol. 111, pp. 7701–7704, 1999.

[103] Popov, A. V., Melvin, J., and Hernandez, R., Dynamics of swelling hard spheres surmised by an irreversible Langevin equation, *J. Phys. Chem. A*, vol. 110, pp. 1635–1644, 2006.

[104] Doering, C. R. and Gadoua, J. C., Resonant activation over a fluctuating barrier, *Phys. Rev. Lett.*, vol. 69, p. 2318, 1992.

[105] Maddox, J., Surmounting fluctuating barriers, *Nature*, vol. 359, p. 771, 1992.

[106] Pechukas, P. and Hänggi, P., Rates of activated processes with fluctuating barriers, *Phys. Rev. Lett.*, vol. 73, p. 2772, 1994.

[107] Reimann, P., Thermally driven escape with fluctuating potentials: A new type of resonant activation, *Phys. Rev. Lett.*, vol. 74, p. 4576, 1995.

[108] Gammaitoni, L., Hänggi, P., Jung, P., and Marchesoni, F., Stochastic resonance, *Rev. Mod. Phys.*, vol. 70, p. 223, 1998.

[109] Shepherd, T. D. and Hernandez, R., Chemical reaction dynamics with stochastic potentials beyond the high-friction limit, *J. Chem. Phys.*, vol. 115, pp. 2430–2438, 2001.

[110] Moix, J. and Hernandez, R., Dissipating the Langevin equation in the presence of an external stochastic potential, *J. Chem. Phys.*, vol. 122, p. 114111, 2005.

[111] Shepherd, T. D. and Hernandez, R., Activated dynamics across stochastic aperiodic potentials, *J. Phys. Chem. B*, vol. 106, pp. 8176–8181, 2002.

[112] Bartsch, T., Hernandez, R., and Uzer, T., Transitions state in a noisy environment, *Phys. Rev. Lett.*, vol. 95, pp. 058301–01–04, 2005.

[113] Bartsch, T., Uzer, T., and Hernandez, R., Stochastic transition states: Reaction geometry amidst noise, *J. Chem. Phys.*, vol. 123, p. 204102, 2005.

[114] Bartsch, T., Uzer, T., Moix, J. M., and Hernandez, R., Identifying reactive trajectories using a moving transition state, *J. Chem. Phys.*, vol. 124, pp. 244310–01–13, 2006.

[115] Zhong, S., Moix, J. M., Quirk, S., and Hernandez, R., Dihedral-angle information entropy as a gauge of secondary structure propensity, *Biophys. J.*, vol. 91, pp. 4014–4023, 2006.

[116] Ozer, G., Foley, J., Zhong, S., Quirk, S., , Moix, J. M., and Hernandez, R. “D_2 check: a web server for protein structure interpretation,” in preparation.

[117] Moix, J. and Hernandez, R. “The role of long-time correlation in dissipative adsorbate dynamics on metal surfaces,” submitted.

[118] Oh, S. M., Kyuno, K., Koh, S. J., and Ehrlich, G., Atomic jumps in surface self-diffusion: W on W(110), *Phys. Rev. B*, vol. 66, p. 233406, 2002.

[119] Backus, E. H. G., Eichler, A., Kleyn, A. W., and Bonn, M., Real-time observation of molecular motion on a surface, *Science*, vol. 310, p. 1790, 2005.

[120] Lide, D. R., ed., *CRC Handbook of Chemistry and Physics.* Boca Raton, FL: CRC Press, 77th ed., 1996.

[121] Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A*, vol. 31, pp. 1695–1697, 1985.

[122] Martyna, G. J., Klein, M. L., and Tuckerman, M., Nosé-Hoover chains: The canonical ensemble via continuous dynamics, *J. Chem. Phys.*, vol. 97, pp. 2635–2645, 1992.

[123] Weakliem, P. C. and Carter, E. A., Constant temperature molecular dynamics simulations of Si(100) and Ge(100): Equilibrium structure and short-time behavior, *J. Chem. Phys.*, vol. 96, pp. 3240–3250, 1992.

[124] Mel’nikov, V. I. and Meshkov, S. V., Theory of activated rate processes: Exact solution of the Kramers Problem, *J. Chem. Phys.*, vol. 85, p. 1018, 1986.

[125] Pollak, E., Grabert, H., and Hänggi, P., Theory of activated rate processes for arbitrary frequency dependent friction: Solution of the turnover problem, *J. Chem. Phys.*, vol. 91, pp. 4073–4087, 1989.

[126] Tully, J. C., Dynamics of Gas surface interactions, *Surf. Sci.*, vol. 111, pp. 461–478, 1982.

[127] Senft, D. C. and Ehrlich, G., Long jumps in surface diffusion: One-dimensional migration of isolated adatoms, *Phys. Rev. Lett.*, vol. 74, p. 294, 1995.

[128] Linderoth, T. R., Horch, S., Lægsgaard, E., Stensgaard, I., and Besenbacher, F., Surface diffusion of Pt on Pt(110); Arrhenius behavior of long jumps, *Phys. Rev. Lett.*, vol. 78, p. 4978, 1997.

[129] Benzi, R., Sutera, A., and Vulpiani, A., The mechanism of stochastic resonance, *J. Phys. A*, vol. 14, p. L453, 1981.

[130] Reimann, P. and Hänggi, P., *Stochastic Dynamics*, vol. 484 of *Lecture Notes in Physics*. Berlin: Springer, 1997.

[131] Astumian, R. and Moss, F., The constructive role of noise in fluctuation driven transport and stochastic resonance, *Chaos*, vol. 8, p. 533, 1998.

[132] Jansch, H., Xu, J., and J.T. Yates, J., Electron stimulated surface migration of CO on Pt(335). First spectroscopic evidence for a new phenomenon, *J. Chem. Phys.*, vol. 99, p. 721, 1993.

[133] Ditchfield, R., Llera-Rodriguez, D., and Seebauer, E., Nonthermal effects of photon illumination on surface diffusion, *Phys. Rev. Lett.*, vol. 81, p. 1259, 1998.

[134] Whitman, L. J., Stroscio, J. A., Dragoset, R. A., and Celotta, R. J., Manipulation of adsorbed atoms and creation of new structures on room-temperature surfaces with a scanning tunneling microscope, *Science*, vol. 251, p. 1206, 1991.

[135] Shklyaev, A. A., Shibata, M., and Ichikawa, M., Formation of Ge nanoislands using a scanning tunneling microscope, *J. Appl. Phys.*, vol. 88, p. 1397, 2000.

[136] Shklyaev, A. A., Shibata, M., and Ichikawa, M., Continuous transfer of Ge by the tip of a scanning tunneling microscope for formation of lines, *J. Vac. Sci. Technol. B*, vol. 19, p. 103, 2001.

[137] Moix, J., Hladik, C., and Hernandez, R. “Is resonant activation a mechanism for patterning adsorbates on a two-dimensional surface?,” in preparation.

[138] Shepherd, T. D. and Hernandez, R., An optimized mean first passage time approach for obtaining rates in activated processes, *J. Chem. Phys.*, vol. 117, pp. 9227–9233, 2002.

[139] Ermak, D. L. and Buckholz, H., Numerical integration of the Langevin equation: Monte Carlo simulation, *J. Comput. Phys.*, vol. 35, pp. 169–182, 1980.

[140] Allen, M. P. and Tildesley, D. J., *Computer Simulations of Liquids.* New York: Oxford, 1987.

[141] Carmeli, B. and Nitzan, A., Theory of activated rate processes: Position dependent friction, *Chem. Phys. Lett.*, vol. 102, p. 517, 1983.

[142] Straub, J. E., Borkovec, M., and Berne, B. J., Molecular dynamics study of an isomerizing diatomic in a Lennard-Jones fluid, *J. Chem. Phys.*, vol. 89, pp. 4833–4847, 1988.

[143] Zhu, S.-B. and Robinson, G. W., Ultrafast dynamics of a quasi-dissociative diatomic molecule in solution, *J. Phys. Chem.*, vol. 93, p. 164, 1988.

[144] Krishnan, R., Singh, S., and Robinson, G. W., Space-dependent friction in the theory of activated rate processes, *Phys. Rev. A*, vol. 45, p. 5408, 1992.

[145] Pollak, E. and Berezhkovskii, A. M., Fokker-Planck equation for nonlinear stochastic dynamics in the presence of space and time dependent friction, *J. Chem. Phys.*, vol. 99, pp. 1344–1346, 1993.

[146] Straus, J. B., Gomez–Llorente, J. M., and Voth, G. A., Manifestations of spatially–dependent friction in classical activated rate processes, *J. Chem. Phys.*, vol. 98, pp. 4082–4097, 1993.

[147] Neria, E. and Karplus, M., A position dependent friction model for solution reactions in the high friction regime: Proton tranfer in triosephosphate isomerase (TIM), *J. Chem. Phys.*, vol. 105, p. 10812, 1996.

[148] Antoniou, D. and Schwartz, S. D., Quantum proton transfer with spatially dependent friction: Phenol- amine in methyl chloride, *J. Chem. Phys.*, vol. 110, p. 7359, 1999.

[149] Haynes, G. R., Voth, G. A., and Pollak, E., A theory for the activated barrier crossing rate constant in systems influenced by space and time dependent friction, *J. Chem. Phys.*, vol. 101, pp. 7811–7822, 1994.

[150] Haynes, G. R., Voth, G. A., and Pollak, E., A theory for the thermally activated rate constant in systems with spatially dependent friction, *Chem. Phys. Lett.*, vol. 207, pp. 309–316, 1993.

[151] Voth, G. A., A theory for treating spatially-dependent friction in classical activated rate processes, *J. Chem. Phys.*, vol. 97, pp. 5908–5910, 1992.

[152] Gertner, B. J., Bergsma, J. P., Wilson, K. R., Lee, S., and Hynes, J. T., Molecular dynamics of a model S_N2 reaction in water, *J. Chem. Phys.*, vol. 86, p. 1377, 1987.

[153] Haynes, G. R. and Voth, G. A., Reaction-coordinate-dependent friction in classical activated barrier crossing dynamics: When it matters and when it doesn't, *J. Chem. Phys.*, vol. 103, pp. 10176–10182, 1995.

[154] Straub, J. E., Borkovec, M., and Berne, B. J., Calculation of dynamic friction in intramolecular freedom, *J. Phys. Chem.*, vol. 91, pp. 4995–4998, 1987.

[155] den Broeck, C. V., Simple stochastic model for resonant activation, *Phys. Rev. E*, vol. 47, p. 4579, 1993.

[156] Reimann, P., Thermally activated escape with potential fluctuations driven by an Ornstein-Ulenbeck process, *Phys. Rev. E*, vol. 52, p. 1579, 1995.

[157] Bartussek, R., Madureira, A. J. R., and Hänggi, P., Surmounting a fluctuating double well: A numerical study, *Phys. Rev. E*, vol. 52, p. R2149, 1995.

[158] Porrà, J. M., Masoliver, J., and Lindenberg, K., Mean exit times for free inertial stochastic processes, *Phys. Rev. E*, vol. 50, p. 1985, 1994.

[159] Talkner, P. and Hänggi, P., eds., *New Trends in Kramers' Reaction Rate Theory*, vol. 11 of *Understanding Chemical Reactivity.* Dordrecht: Kluwer Academic Publishers, 1995.

[160] Miller, W. H., Spiers Memorial Lecture Quantum and semiclassical theory of chemical reaction rates, *Faraday Disc. Chem. Soc.*, vol. 110, p. 1, 1998.

[161] G.Truhlar, D., Garrett, B. C., and Klippenstein, S. J., Current status of transition-state theory, *J. Phys. Chem.*, vol. 100, p. 12771, 1996.

[162] Jaffé, C., Ross, S. D., Lo, M. W., Marsden, J. E., Farrelly, D., and Uzer, T., Statistical theory of asteroid escape rates, *Phys. Rev. Lett.*, vol. 89, p. 11101, 2002.

[163] Koon, W. S., Lo, M. W., Marsden, J. E., and Ross, S. D., Heteroclinic connections between periodic orbits and resonance transitions in celestial mechanics, *Chaos*, vol. 10, no. 2, pp. 427–469, 2000.

[164] Komatsuzaki, T. and Berry, R. S., Chemical Reaction Dynamics: Many-Body Chaos and Regularity, *Adv. Chem. Phys.*, vol. 123, pp. 79–152, 2002.

[165] C. Jaffé, D. Farrelly, and T. Uzer, Phys. Rev. Lett. **84**, 610 (2000); Phys. Rev. A **60**, 3833 (1999).

[166] Eckhardt, B., Transition state theory for ballistic electrons, *J. Phys. A*, vol. 28, p. 3469, 1995.

[167] Toller, M., Jacucci, G., DeLorenzi, G., and Flynn, C. P., Theory of classical diffusion jumps in solids, *Phys. Rev. B*, vol. 32, p. 2082, 1985.

[168] Truhlar, D. G., Issacson, A. D., and Garrett, B. C., Theory of Chemical Reaction Dynamics, vol. 4, pp. 65–137, Boca Raton, FL: CRC, 1985.

[169] Grote, R. F. and Hynes, J. T., The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models, *J. Chem. Phys.*, vol. 73, pp. 2715–32, 1980.

[170] Truhlar, D. G. and Garrett, B. C., Variational Transition State Theory, *Annu. Rev. Phys. Chem.*, vol. 35, pp. 159–189, 1984.

[171] Pollak, E., Variational transition state theory for activated rate processes, *J. Chem. Phys.*, vol. 93, pp. 1116–1124, 1990.

[172] Pollak, E., Canonical variational transition state theory for dissipative systems: Application to generalized Langevin equations, *J. Chem. Phys.*, vol. 96, pp. 8877–8888, 1992.

[173] Tucker, S. C. and Pollak, E., Microcanonical variational transition-state theory for reaction rates in dissipative systems, *J. Stat. Phys.*, vol. 66, pp. 975–90, 1992.

[174] Tucker, S. C., Variational transition state theory in condensed phases, in *New Trends in Kramers' Reaction Rate Theory* (Hänggi, P. and Talkner, P., eds.), pp. 5–46, The Netherlands: Kluwer Academic, 1995.

[175] Straub, J. E., Borkovec, M., and Berne, B. J., Shortcomings of current theories of mon-Markovian activated rate processes, *J. Chem. Phys.*, vol. 83, pp. 3172–4, 1985.

[176] Straub, J. E., Borkovec, M., and Berne, B. J., Non-Markovian activated rate processes: Comparison of current theories with numerical simulation data, *J. Chem. Phys.*, vol. 84, pp. 1788–94, 1986.

[177] Pollak, E. and Talkner, P., Transition-state recrossing dynamics in activated rate processes, *Phys. Rev. E*, vol. 51, p. 1868, 1995.

[178] Hynes, J. T., The theory of reactions in solution, in *Theory of Chemical Reaction Dynamics* (Baer, M., ed.), vol. 4, pp. 171–234, Boca Raton, FL: CRC, 1985.

[179] Pollak, E., Theory of activated rate processes: A new derivation of Kramers' expression, *J. Chem. Phys.*, vol. 85, pp. 865–867, 1986.

[180] Zwanzig, R., Nonlinear generalized Langevin equation, *J. Stat. Phys.*, vol. 9, p. 215, 1973.

[181] Caldeira, A. O. and Leggett, A. J., Influence of dissipation on quantum tunneling in macroscopic systems, *Phys. Rev. Lett.*, vol. 46, p. 211, 1981. Ann. Phys. (N. Y.) **149**, 374 (1983).

[182] Graham, R., Macroscopic Theory of Activated Decay of Metastable States, *J. Stat. Phys.*, vol. 60, no. 5/6, pp. 675–694, 1990.

[183] Bartsch, T., Hernandez, R., and Uzer, T., Transition state in a noisy environment, *Phys. Rev. Lett.*, vol. 95, p. 058301, 2005.

[184] Bartsch, T., Uzer, T., and Hernandez, R., Stochastic transition states: Reaction geometry amidst noise, *J. Chem. Phys.*, vol. 123, p. 204102, 2005.

[185] Martens, C. C., Qualitative dynamics of generalized Langevin equations and the theory of chemical reaction rates, *J. Chem. Phys.*, vol. 116, no. 6, pp. 2516–2528, 2002.

[186] Hernandez, R., A combined use of perturbation theory and diagonalization: Application to bound energy levels and semiclassical rate theory, *J. Chem. Phys.*, vol. 101, pp. 9534–9547, 1994.

[187] Keck, J. C., Statistical Investigation of Dissociation Cross-Sections for Diatoms, *Discuss. Faraday Soc.*, vol. 33, pp. 173–182, 1962.

[188] Keck, J. C., Variational Theory of Reaction Rates, *Adv. Chem. Phys.*, vol. 13, pp. 85–121, 1967.

[189] Grimmelmann, E. K., Tully, J. C., and Helfand, E., Molecular dynamics of infrequent events: Thermal desorption of xenon from a platinum surface, *J. Chem. Phys.*, vol. 74, p. 5300, 1981.

[190] Rejto, P. A. and Andersen, H. C., A reactive-flux theory of chemical surface diffusion, *J. Chem. Phys.*, vol. 92, p. 6217, 1990.

[191] Dellago, C., Bolhuis, P., Csajka, F. S., and Chandler, D., Transition path sampling and the calculation of rate constants, *J. Chem. Phys.*, vol. 108, p. 1964, 1998.

[192] Dellago, C., Bolhuis, P., and Chandler, D., On the calculation of rate constants in the transition path ensemble, *J. Chem. Phys.*, vol. 110, p. 6617, 1999.

[193] Bolhuis, P. G., Chandler, D., Dellago, C., and Geissler, P., Transition Path Sampling: Throwing ropes over mountain passes, in the dark, *Annu. Rev. Phys. Chem.*, vol. 53, pp. 291–318, 2002.

[194] Dellago, C., Bolhuis, P. G., and Geissler, P., Transition Path Sampling, *Adv. Chem. Phys.*, vol. 123, pp. 1–78, 2002.

[195] MacFadyen, J. and Andricioaei, I., A skewed-moment method to efficiently generate conformational-transition trajectories, *J. Chem. Phys.*, vol. 123, p. 074107, 2005.

[196] Yamamoto, T., Quantum Statistical Mechanical Theory of the Rate of Exchange Chemical Reactions in the Gas Phase, *J. Chem. Phys.*, vol. 33, pp. 281–289, 1960.

[197] Zwanzig, R., Time-correlation functions and transport coefficients in statistical mechanics, *Annu. Rev. Phys. Chem.*, vol. 16, p. 67, 1965.

[198] Xing, C. and Andricioaei, I., On the calculation of time correlation functions by potential scaling, *J. Chem. Phys.*, vol. 123, p. 034110, 2006.

[199] Abramowitz, M. and Stegun, I. A., *Pocketbook of Mathematical Functions*. Frankfurt/Main: Verlag Harri Deutsch, 1984.

[200] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E., The Protein Data Bank, *Nucleic Acids Research*, vol. 28, pp. 235–242, 2000.

[201] Branden, C. I. and Jones, T. A., Between objectivity and subjectivity, *Nature*, vol. 343, pp. 687–689, 1990.

[202] Hooft, R. W. W., Vriend, G., Sander, C., and Abola, E. E., Errors in protein structures, *Nature*, vol. 381, p. 272, 1996.

[203] Abola, E. E., Bairoch, A., Barker, W. C., Beck, S., Benson, D. A., Berman, H., Cameron, G., Cantor, C., Doubet, S., Hubbard, T. J. P., Jones, T. A., Kleywegt, G. J., Kolaskar, A. S., Van Kuik, A., Lesk, A. M., Mewes, H. W., Neuhaus, D., Pfeiffer, G., TenEyck, L. F., Simpson, R. J., Stoesser, G., Sussman, J. L., Tateno, Y., Tsugita, A., Ulrich, E. L., and Vliegenthart, J. F. G., Quality control in databanks for molecular biology, *BioEssays*, vol. 22, pp. 1024–1034, 2000.

[204] Ramakrishnan, C. and Ramachandran, G. N., Stereochemical criteria for polypeptide and protein chain conformations II. allowed conformations for a pair of peptide units, *Biophys. J.*, vol. 5, pp. 909–933, 1965.

[205] Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M., Stereochemical quality of protein structure coordinates, *Proteins*, vol. 12, pp. 345–364, 1992.

[206] Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Cryst.*, vol. 26, pp. 283–291, 1993.

[207] MacArthur, M. W. and Thornton, J. M., Conformation analysis of protein structures derived from NMR data, *Proteins*, vol. 17, pp. 232–251, 1993.

[208] MacArthur, M. W., Laskowski, R. A., and Thornton, J. M., Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy, *Curr. Opin. Struct. Biol.*, vol. 4, pp. 731–737, 1994.

[209] Laskowski, R. A., MacArthur, M. W., and Thornton, J. M., Validation of protein models derived from experiment, *Curr. Opin. Struct. Biol.*, vol. 8, pp. 631–639, 1998.

[210] Brünger, A. T., Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature*, vol. 355, pp. 472–475, 1992.

[211] Kleywegt, G. J. and Jones, T. A., Phi/Psi-chology: Ramachandran revisited, *Structure*, vol. 4, pp. 1395–1400, 1996.

[212] Kleywegt, G. J., Validation of protein models from C^{α} coordinates alone, *J. Mol. Biol.*, vol. 273, pp. 371–376, 1997.

[213] Kleywegt, G. J. and Jones, T. A., Model building and refinement practice, *Methods. Enzymol.*, vol. 277, pp. 208–230, 1997.

[214] Kleywegt, G. J., Validation of protein crystal structures, *Acta Cryst.*, vol. D56, pp. 249–265, 2000.

[215] Hooft, R. W. W., Sander, C., and Vriend, G., Objectively judging the quality of a protein structure from a Ramachandran plot, *CABIOS*, vol. 13, pp. 425–430, 1997.

[216] Lovell, S. C., Davis, I. W., Arendall III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C., Structure validation by C_α geometry: ϕ, ψ and C_β deviation, *Proteins*, vol. 50, pp. 437–450, 2003.

[217] Willard, L., Ranjan, A., Zhang, H. Y., Monzavi, H., Boyko, R. F., Sykes, B. D., and Wishart, D. S., VADAR: a web server for quantitative evaluation of protein structure quality, *Nucleic Acids Research*, vol. 31, pp. 3316–3319, 2003.

[218] Sudarsanam, S., DuBose, R. F., March, C. J., and Srinivasan, S., Modeling protein loops using a ϕ_{i+1},ψ_i dimer database, *Protein Sci.*, vol. 4, pp. 1412–1420, 1995.

[219] Sudarsanam, S. and Srinivasan, S., Searching for protein loops in parallel, *CABIOS*, vol. 11, pp. 591–593, 1995.

[220] Sudarsanam, S. and Srinivasan, S., Sequence-dependent conformational sampling using a database of ϕ_{i+1} and ψ_i angles for predicting polypeptide backbone conformations, *Protein Eng.*, vol. 10, pp. 1155–1162, 1997.

[221] Parker, J. M. R., The relationship between peptide plane rotation (PPR) and similar conformations, *J. Comp. Chem.*, vol. 20, pp. 947–955, 1999.

[222] Shortle, D., Composite of local structure propensities: evidence for local encoding of long-rang structure, *Protein Sci.*, vol. 11, pp. 18–26, 2002.

[223] Shortle, D., Propensities, probabilities, and the Boltzmann hypothesis, *Protein Sci.*, vol. 12, pp. 1298–1302, 2003.

[224] Fang, Q. J. and Shortle, D., A Consistent set of statistical potentials for quantifying local side-chain and backbone interactions, *Protein Sci.*, vol. 60, pp. 90–96, 2005.

[225] Fang, Q. J. and Shortle, D., Enhanced Sampling near the Native Conformation Using Statistical Potentials for Local Side-Chain and Backbone Interactions, *Proteins*, vol. 60, pp. 97–102, 2005.

[226] Hovmöller, S., Zhou, T., and Ohlson, T., Conformations of amino acids in proteins, *Acta Cryst.*, vol. D58, pp. 768–776, 2002.

[227] Sheik, S. S., Ananthalakshmi, P., Bhargavi, G. R., and Sekar, K., CADB: Conformation angles dataBase of proteins, *Nucleic Acids Research*, vol. 31, pp. 448–451, 2003.

[228] Priestle, J. P., Improved dihedral-angle restraints for protein structure refinement, *J. Appl. Cryst.*, vol. 36, pp. 34–42, 2003.

[229] Dayalan, S., Bevinakoppa, S., and Schroder, H., A dihedral angle database of short sub-sequences for protein structure prediction, The Second Asia-Pacific Bioinfomatics Conference, Australian Computer Society, INC., 2004.

[230] Vriend, G., WHAT IF: a molecular modeling and drug design program, *J. Mol. Graph.*, vol. 8, pp. 52–55, 1990.

[231] Zheng, Q., Rosenfeld, R., DeLisi, C., and Kyle, D. J., Multiple copy sampling in protein loop modeling: computational efficiency and sensitivity to dihedral angle perturbations, *Protein Sci.*, vol. 3, pp. 493–506, 1994.

[232] Mathiowetz, A. M. and Goddard III, W. M., Building proteins from C_α coordinates using the dihedral probability grid Monte Carlo method, *Protein Sci.*, vol. 4, pp. 1217–1232, 1995.

[233] Cheng, B., Nayeem, A., and Scheraga, H. A., From secondary structure to three-dimensional structure:improved dihedral angle probability distribution function for use with energy searches for native structures of polypeptides and proteins, *J. Comp. Chem.*, vol. 17, pp. 1453–1480, 1996.

[234] Fiser, A., Gian Do, R. K., and Šali, A., Modeling of loops in protein structures, *Protein Sci.*, vol. 9, pp. 1753–1773, 2000.

[235] Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Šali, A., Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct.*, vol. 29, pp. 291–325, 2000.

[236] Baker, D. and Šali, A., Protein structure prediction and structural genomics, *Science*, vol. 294, pp. 93–96, 2001.

[237] Fiser, A., Feig, M., Brooks III, C. L., and Šali, A., Evolution and physics in comparative protein structure modeling, *Acc. Chem. Res.*, vol. 35, pp. 413–421, 2002.

[238] Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., and Friesner, R. A., A hierarchical approach to all-atom protein loop prediction, *Proteins*, vol. 55, pp. 351–367, 2004.

[239] Wu, T. T. and Kabat, E. A., An attempt to locate the non-helical and permissively helical sequences of proteins: application to the variable regions of immunoglobulin light and heavy chains, *Proc. Natl. Acad. Sci. USA*, vol. 68, pp. 1501–1506, 1971.

[240] Kabat, E. A. and Wu, T. T., Construction of a three-dimensional model of the polypeptide backbone of the variable region of kappa immunoglobulin light chains, *Proc. Natl. Acad. Sci. USA*, vol. 69, pp. 960–964, 1972.

[241] Wu, T. T. and Kabat, E. A., Attempt to evaluate influence of neighboring amino-acid (n-1) and (n+1) on backbone conformation of amino-acid (n) in proteins-use in predicting 3-dimensional structure of polypeptide backbone of other proteins, *J. Mol. Biol.*, vol. 75, pp. 13–31, 1973.

[242] Pappu, R. V., Srinivasan, R., and Rose, G. D., The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding, *Proc. Natl. Acad. Sci. USA*, vol. 97, pp. 12565–12570, 2000.

[243] Chakrabarti, P. and Pal, D., The interrelationships of side-chain and main-chain conformations in proteins, *Prog. Biophys. Mol. Biol.*, vol. 76, pp. 1–102, 2001.

[244] Zaman, M. H., Shen, M. Y., Berry, R. S., Freed, K. F., and Sosnick, T. R., Investigations into sequence and conformational dependence of backbone entropy,inter-basin dynamics and the Flory isolated-pair hypothesis for peptides, *J. Mol. Biol.*, vol. 331, pp. 693–711, 2003.

[245] Betancourt, M. R. and Skolnick, J., Local propensities and statistical potentials of backbone dihedral angles in proteins, *J. Mol. Biol.*, vol. 342, pp. 635–649, 2004.

[246] Esposito, L., De Simone, A., Zagari, A., and Vitagliano, L., Correlation between ω and ψ dihedral angles in protein structures, *J. Mol. Biol.*, vol. 347, pp. 483–487, 2005.

[247] RCSB Protein Data Bank. `http://www.rcsb.org/pdb/clusterStatistics.do`, April, 2005.

[248] DeWitte, R. S. and Shakhnovich, E. I., Pseudodihedrals: Simplified protein backbone representation with knowledge-based energy, *Protein Sci.*, vol. 3, pp. 1570–1581, 1994.

[249] Shannon, C. E., A Mathematical Theory of Communication, *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

[250] Solis, A. D. and Rackovsky, S., Optimally informative backbone structural propensities in proteins, *Proteins*, vol. 48, pp. 463–486, 2002.

[251] Solis, A. D. and Rackovsky, S., On the use of secondary structure in protein structure prediction: a bioinformatic analysis, *Polymer*, vol. 45, pp. 525–546, 2004.

[252] Garnier, J., Osguthorpeb, D. J., and Robson, B., Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins, *J. Mol. Biol.*, vol. 120, pp. 97–120, 1978.

[253] Kloczkowski, A., Ting, K. L., Jernigan, R. L., and Garnier, J., Combining the GOR algorithm with evolutionary information for protein secondary structure prediction from amino acid sequence, *Proteins*, vol. 49, pp. 154–166, 2002.

[254] Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A., Structure of HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody, *Nature*, vol. 393, pp. 648–659, 1998.

[255] Cho, Y. J., Gorina, S., Jeffrey, P. D., and Pavletich, N. P., Crystal-structure of a p53 tumor suppressor DNA complex- understanding tumorigenic mutations, *Science*, vol. 265, pp. 346–355, 1994.

[256] Wall, M. A., Coleman, D. E., Lee, E., Iniguez-Lluhi, J. A., Posner, B. A., Gilman, A. G., and Sprang, S. R., The structure of the G-protein heterotrimer G(I-alpha-1)beta(1)gamma(2), *Cell*, vol. 83, pp. 1047–1058, 1995.

[257] Sali, A. and Blundell, T. L., Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.*, vol. 234, pp. 779–815, 1993.

[258] Zhong, S., Quirk, S., and Hernandez, R. “A checking function for secondary structure propensity of proteins at the residute level,” in preparation.

[259] Hynes, T. R. and Fox, R. O., The crystal structure of staphylococcal nuclease refined at 1.7 angstrom resolution, *Proteins*, vol. 10, pp. 92–105, 1991.

[260] Zhang, Y., Hubner, I. A., Arakaki, A. K., Shakhnovich, E., and Skolnick, J., On the origin and highly likely completeness of single-domain protein structures, *Proc. Natl. Acad. Sci. USA*, vol. 103, pp. 2605–2610, 2006.

[261] Flory, P. J., *Statistical Mechanics of Chain Molecules.* New York: Wiley-Interscience, 1969.

[262] Ohkubo, Y. Z. and Brooks III, C. L., Exploring Flory’s isolated-pair hypothesis: statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A, *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 13916–13921, 2003.