Title
Multi-scale simulations of biophysics problems, with emphasis on Importance Sampling

Author
Lin, Chen

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Multi-scale simulations of biophysics problems, with emphasis on Importance Sampling

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Chemistry

by

Chen Lin

2021

ABSTRACT OF THE DISSERTATION

Multi-scale simulations of biophysics problems, with emphasis on Importance Sampling

by

Chen Lin
Doctor of Philosophy in Chemistry
University of California, Los Angeles, 2021
Professor Robijn F. Bruinsma, Chair

The behavior of a biophysical system often is quite different when investigated on different length scales and a complete description of such a system typically requires different approaches for these different length scales. At atomic and mesoscopic scales, a considerable number of degrees of freedom are involved. The associated free energy profiles are relatively rugged with energy barriers or kinetic bottlenecks preventing efficient sampling when doing numerical simulations. To alleviate this problem, a variety of sampling strategies have been developed.

In this thesis, I will start with general backgrounds of multi-scale modeling of biophysical systems, and review the different sampling strategies that are being used when doing simulations. Next, I will talk about three of my projects. The first two involve protein binding problems at both the microscopic and the mesoscopic level. The third one is the diffusion-driven phase separation problem inside a cross-linked network of semiflexible polymers. I will use these three projects to illustrate how one can apply multi-scale descriptions and sampling techniques to biophysical systems in and out of equilibrium.

The dissertation of Chen Lin is approved.

Alexander J. Levine

William M. Gelbart

Justin R. Caram

Robijn F. Bruinsma, Committee Chair

University of California, Los Angeles

2021

To the landscapes through the journey

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It has been a long journey, way longer than one can imagine. I first want to thank my undergraduate advisor, professor Joseph Francisco. Without the experience I had at "NC3", Nebraska, I would not have been so enlightened about the numerical simulation of chemical and biophysical systems. That experience switched my mindset from small molecular optimizations at the quantum mechanical level to the simulation of dynamics of a many-body system.

In my first year of Ph.D., I had the pleasure of joining professor Robijn Bruinsma's group. Thanks to his patience and broad knowledge, I reshaped my understanding of all-atoms simulation and enhanced sampling techniques to a physics' point of view at the early stage of our research. His willingness to engage in different research topics also conferred me a flexibility to sculpture my own interests. After working with him for five years, I systematically learned how to describe a system as a mathematical model beyond all atoms level and truncate it with both analytical and numerical strategies. This fashion of doing research will facilitate me for my future work. I would also appreciate the passionate carefulness from Robijn for my personal life, especially during tough circumstances like the ongoing pandemic.

Professor Alex Levine and professor Joseph Rudnick also strongly influenced my research and studies. Studying Alex's mean-field statistical mechanics and being his teaching assistant in a class on particle statistical mechanics enhanced my understanding of this area. Our collaborated work focusing on a phase separation problem further pushed me to pursue the knowledge of systems under nonequilibrium conditions. Joe, on the other hand, provided me more insight from a mathematical perspective. His flair for solving mathematical models with complicated boundary conditions and truncate equations in different ways solidified my understanding of modeling.

I also warmly thank my committee members, professor William Gelbart and professor Justin Caram. Without Bill's effort in helping me find an advisor, all the above would have been impossible. Bill and Justin were always patient and enthusiastic when discussing my

research and career plans. I also specially thank professor Chen Yang from Boston College, who kindly offered me her opinion on the balance between the industrial and academic worlds. I greatly appreciate their insightful comments, advice and encouragement.

My group mates and friends played a major role in my study at UCLA. I had the pleasure of working closely with Zach Gvildys. We learned from each other a lot especially when facing profound mathematical approaches, efficient discussions with him speed up my research progress considerably. It was always nice to have someone else to discuss interesting problems with, for which I greatly appreciate Dr. Jonathan Kernes, Baochen She, Peiqi Wang and Xiang Yuan who walked with me on the long road of learning fundamental physics. I thank their enthusiasm and curiosity when figuring out tough problems with me, even out of their research area. Shaoqing Ai and Chenkai Wang, on the other hand, have provided me help technically. Shaoqing taught me how to programming from 0 to 1; Chenkai offered me a deeper understanding of probability theory and stochastic process. Without them, I would not have achieved anything.

I would also thank my basketball teammates, who kept me entertained in my life outside of scientific studies. The bronze medal we won in spring 2019 will always be a golden memory. I also had great times with Ruxi Dai, Tian Deng, Rong Duan, Yao Gong, Mingxuan Lu, Yuting Miao, Zhe Wang, Yixin Wu, and Guangyan Zhong, I am grateful to thank their encouragement and accompany during the pandemic.

I would like to thank the NSF-DMR for support under CMMT Grant No.1836404.

Finally, I would like to thank my mom for her unconditional support, who always encouraged me to pursue the area of my interests both in science and basketball.

VITA

2013–2016 B.S. in Chemistry, Purdue University, West Lafayette, IN

2016–2021 Ph.D. in Chemistry, UCLA, Los Angeles, CA

PUBLICATIONS

Fluid Droplet growth in a cross-linked network. ***in preparation.***
Chen Lin, Alex Levine, Robijn Bruinsma.

Theory of Diffusion-Limited Reactions between Proteins in a Confined Environment. ***in preparation.***
Chen Lin, Zach Gvildys, Joseph Rudnick, Robijn Bruinsma.

Specific inter-domain interactions stabilize a compact HIV-1 Gag conformation. ***PloS one*****, 14(8).**
Chen Lin, Paola Mendoza-Espinosa, Ioulia Rouzina, Orlando Guzmán, José Antonio Moreno-Razo, Joseph S Francisco, Robijn Bruinsma.

Interaction of the ethyl radical with a water droplet. ***in preparation.***
Chen Lin, Ivan Gladich, Chongqin Zhu, Jie Zhong, Joseph S. Francisco.

Intramolecular hydrogen bonding in malonaldehyde and its radical analogues. ***J. Chem. Phys.*** **2017, 147(12): 124309.**
Chen Lin, Manoj Kumar, Brian A Finney, Joseph S Francisco.

Spectroscopic characterization of the ethyl radical-water complex. ***J. Chem. Phys.2016, 145(14): 144301.***

Chen Lin, Brian A Finney, Allan H Laufer, Josep M Anglada, Joseph S Francisco.

CHAPTER 1

Introduction

1.1 Multi-scale modeling of problems in biophysics.

Multi-scale modeling describes research projects that focus on features of a system at multiple scales of time or space. Important problems such as the physics of polymers [2], proteins, DNA, RNA and many other biological macro-molecules [3] have been approached in this manner.

The behavior of a system can be very different when investigated in different scales. On each scale level, particular approaches are used for the description of a system. For example in biophysics, many phenomena can be approached at four different levels. At the quantum mechanical level, electron wavefunctions are computed. Chemical reaction kinetics involving bond breaking and reformation, including catalytic processes, can be studied at this level [4]. Next, at a "quasi-microscopic" scale, chemical bonds are simplified as effective potentials and molecular dynamics methods are used to simulate the dynamics by solving numerically *classical* Newton equations of motion [5]. The forces between the molecules are calculated using interatomic potentials & force fields from a large library, such as GROMACS [6], CHARMM [7] and AMBER [8], that was computed separately, when necessary at the quantum level. These force fields usually have five terms [9]:

$$E_{tot} = E_{stretch} + E_{angle} + E_{dihedral} + E_{electro} + E_{VdW} \tag{1.1.1}$$

where the red terms represent interactions of permanent bonds, such as covalent bonds. These often are represented by effective harmonic springs. The blue terms are reversible, non-

bonding interactions, such as van der Waals interactions and hydrogen bonding. Examples of studies at this level are detailed investigations of the mechanisms of protein conformation transfer and protein folding [10].

Next, on the "mesoscale" level, clusters of groups of atoms or molecules are considered as large quasi-molecules ("coarse-graining"). The effects of interactions with the solvent are included as a combination of friction or viscosity and random noise. The researched object typically is performing a form of Brownian motion due to collisions with the solvent molecules that represent thermal fluctuations. At this scale, the slow diffusive dynamics of large molecules can be studied, where the detailed electronic structure and molecular conformation changes are coarse grained as boundary conditions [11]. Finally, at the macro-scale level the system is investigated by a combination of continuum mechanics and thermodynamics. At this scale, thermal fluctuations are ignored [12].

The first study in this thesis focuses on the intra-molecular binding of the very large *Gag protein*, the structural protein of the HIV-1 virus. The intramolecular binding mechanism was first studied at the microscopic level by all-atom MD simulations. While this gave important information on the binding mechanism, it could not handle the long time scales associated with the diffusive binding/unbinding kinetics. In order to go beyond the limitations of this approach, we developed a coarse-grained model at the mesoscale level using Brownian dynamics simulations and applied methods of non-equilibrium thermodynamics. Finally, we studied the problem of phase separation of proteins at a coarse-grained level to gain a better understanding of the non-equilibrium aspects.

1.2 Enhanced sampling and collective variables (CV).

All-atom simulations of complex systems by *Molecular Dynamics (MD)* simulations is a powerful tool in many branches of physical science. However, for very large biological macromolecules relaxation times scales may range from microseconds up to seconds. For typical MD simulation time step in the femtosecond to picosecond range [13], this is far beyond our

current computational resources. One important reason for these long time scales is that the relaxation may depend on *rare events* where the system passes between different sectors of configuration space separated by high free energy barriers. Kinetic traps in the form of long-lived metastable states may be another reason for long relaxation times.

To alleviate these problems, the computational hardware has been improved, such as the use of GPUs [14]. Separately, new software has been developed in the form of *enhanced sampling techniques*. These sampling techniques can be categorized into three classes: *tempering techniques* [15], *collective variable (CV) based enhanced sampling* [13] [16] and *mathematical approaches* [17]. The first one is beyond the scope of our research while we defer discussing new mathematical methods to a later section.

The key idea of the CV-based sampling method is to integrate over the "fast" degrees of freedom while fixing a limited number of "slow" degrees of freedom (DoF), known as collective variables (CVs). Collective variables are in general functions of the microscopic configuration that have to be integrated over while keeping the CVs fixed. They may have a simple geometric interpretation, such as the separation between two subgroups of a macromolecule or torsion angles when two subgroups are rotated with respect to each other. The CVs also may be large amplitude versions of the slowest relaxation modes of the groundstate. Other choices for the CVs may be the total potential energy or entropy of the system. When simulating a phase transition, the Landau order parameter is the natural choice for the CV. The proper choice of the CVs is a central question for enhanced sampling computational methods [13] [16].

Assume that a CV, or set of CVs, denoted by s has been identified. The probability that the CV adopts a certain value under conditions of thermal equilibrium is given by:

$$P(s) = \frac{\int d^N r \delta[s(r) - s)] e^{-\beta U(r)}}{\int d^N r e^{-\beta U(r)}} \tag{1.2.1}$$

Here, U is the potential energy for the micro-configurations with N the total number of DoF. More precisely, $P(s)ds$ is the probability of finding the CV in the small interval ds of CV

variables. The free energy difference between two sets of CV values i and j is

$$\Delta Q(\mathbf{s}_{ij}) = -\frac{1}{\beta} ln \frac{P(\mathbf{s_i})}{P(\mathbf{s_j})}.$$

If a reference point is defined, one can sketch a free energy surface with respect to that point, which is known as the "potential of mean force".

As mentioned, calculating $Q(s_{ij})$ becomes very difficult when phase space is divided into sectors separated by large energy barriers. To speed up the calculation, one can add a *bias potential* V(**s**) to the energy of the system:

$$P(s)_b = \frac{\int d^N \mathbf{r} \delta[s(\mathbf{r}) - \mathbf{s})] e^{-\beta[U(\mathbf{r})+V(\mathbf{s}(\mathbf{r}))]}}{\int d^N r e^{-\beta[U(\mathbf{r})+V(s(\mathbf{r}))]}} \tag{1.2.2}$$

This bias potential will flatten the free energy surface and push the simulation towards rare regions along selected CV. Finally, one needs to reconstruct the real distribution from the biased simulation data:

$$P(\mathbf{s}) = P(\mathbf{s})_b e^{\beta V(\mathbf{s})} < e^{-\beta V(\mathbf{s}(\mathbf{r}))} > \tag{1.2.3}$$

A number of methods belonging to this class have been developed, The Parrinello group contributed a lot of them. A popular approach is *Metadynamics*, a method that depends on adding a history dependent biased potential. This potential can be regarded as a set of Gaussians deposited along the system trajectory in the selected CVs space to discourage the system from revisiting configurations that already been sampled and force it to oversample rare but events such as barrier crossing [13][18]. A related approach is *Variationally Enhanced Sampling (VES)* [19] that fills simultaneously gaps in the whole free energy surface by guessing the bias potential in the form of a linear expansion of basis functions. The coefficients of the expansion are obtained by minimization of a functional of the bias potential. The minimization process changes the set of coefficient of expansion to adjust the shape of the bias potential to match the gaps in free energy surface. The form of functional was inspired by Kullback-Leibler divergence of information theory [20]. It actually maximizes the relative entropy during the optimization.

Other methods are adaptive force bias [21], steered MD [22] and local elevation [23]. They are all aimed at increasing the accuracy of the selection of bias potential and the reconstruction of probability distribution in different ways. A useful review is ref [24]. In this thesis *Umbrella Sampling* [25] [16] is used, developed by Torrie and Valleau. The sampling is implemented by many windows along the selected CV, with each of them confined by a bias harmonic potential. Thus the sampling is enhanced within a certain region. Then one uses the *Weighted Histogram Analysis Method (WHAM)* [26] to reconstruct the biased simulation data by collecting information from all windows. In the following section we apply the umbrella sampling method combined with WHAM to the Gag protein.

CHAPTER 2

The Gag Protein Problem

2.1 Introduction of the Gag protein problem

Despite intense research efforts, important aspects of the life-cycle of the HIV-1 virus are not well understood [27]. One of these unresolved issues concerns the initiation of the assembly of the protein shell, or capsid, of HIV-1 virus particles (or virions). This capsid is composed of the very large Gag polyproteins (from group-specific antigen). It encloses the viral genome composed of two 10 kilobase single-stranded RNA molecules, to be denoted by gRNA. During assembly, the gRNA molecules must be selected from among an overwhelming majority of host mRNA material present in the cytoplasm. Gag protein is the only protein that is required for the production of virus-like particles (VLPs). Expression of just Gag in non-infected cells – so with no gRNA molecules present – leads to the release of non-infective VLPs that are morphologically indistinguishable from infectious viruses but that package host mRNA molecules instead of gRNA. The unresolved issue in this context concerns the gRNA selection mechanism: despite intense in-vivo and in-vitro studies by many labs, HIV-1 Gag proteins appear to have hardly any binding specificity for gRNA over generic RNA. Yet, over 90 percent of the virions produced in the infected cell do carry HIV-1 gRNA molecules. RNA selection is known to take place during the very first assembly steps, which involves at most only few Gag molecules, so this strange selection mechanism must be a feature of either individual Gag proteins or of a small number of interacting Gag proteins.

As shown in Fig.1, the Gag polyprotein is composed of a number of domains – connected by unstructured, flexible linkers – that are conventionally labelled as MA, CA-NTD, CA-CTD and NC (MA stands for "membrane associated", CA stands for "capsid", NTD and

CTD refer to the N and C terminals of the CA domain, while NC stands for "nucleo-capsid"). In the assembled state, shown in Fig.1, the Gag proteins form a hexagonal lattice [28]

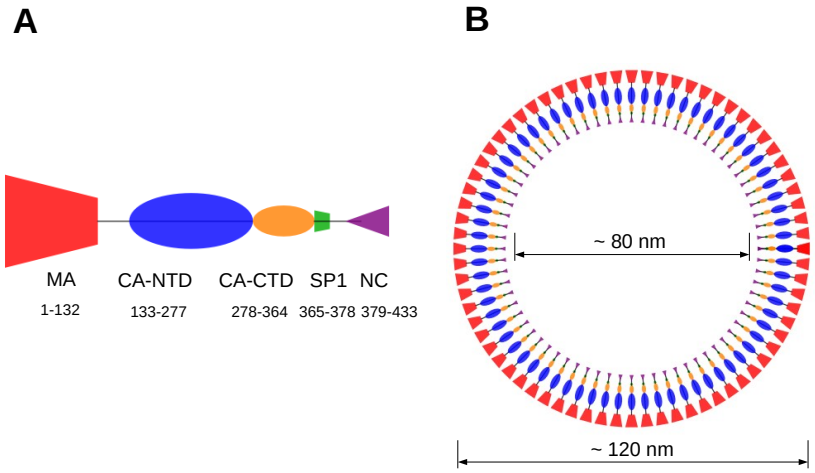


Figure 2.1.1: (A) Schematic of the domains of Gag. The MA domain has a large number of positively charged residues (at neutral pH) while the CA domain, which is composed of the linked CA-NTD and CA-CTD subdomains, is close to neutral. The NC (nucleocapsid) domain, which is separated from the CA-CTD subdomain by the short SP1 sequence, has a net positive charge. (B) Schematic of the immature capsid. The positively charged MA domains of the Gags are associated with the negatively charged plasma membrane (PM). The CA-CTD domains of adjacent Gags are bonded by hydrophobic interactions while positively charged NC domains are associated with the negatively charged viral RNA molecules (not shown).

stabilized (in part) by hydrophobic interactions between the CA-CTD domains. In solution Gag proteins largely are monomeric. A single-molecule FRET study (fluorescent recovery after photobleaching) of Gag proteins in solution [29] reported that they have a range of conformations, some with the MA and NC domains in close proximity (to be denoted by "F-Gag") and some with the two domains far from each other (to be denoted by "U-Gag"). The

U-Gag state has a broad MA-NC distance probability distribution with relatively large MA-NC distances, suggesting that the U-Gag state has significant conformational fluctuations. In contrast, the F-Gag state has a narrow MA-NC distance distribution with relatively short MA-NC distances. The F-Gag and U-Gag states of Gag seem to be stable and do not interconvert over typical FRET observation times (up to 100 s). Importantly, the F-state is the majority component of monomeric Gag in solution. In a recent MD simulation of Gag [30], the different domains of Gag were found to move as rigid solid bodies linked by flexible tethers if the initial state corresponded to the unfolded U-Gag state of the FRET experiments. The MA and NC domains are not correlated in this state. The authors also identified a folded state, with the MA and CA-CTD domains in close contact with each other. In this state, which could correspond to the F-state, the MA and NC domains are quite correlated, indicative of long-range allosteric interactions [31]. Both states remained stable over a simulation time of 300 ns.

The fact that the majority component of Gag in solution is in the F-Gag state suggests a possible gRNA selection mechanism. Assume that the observed lack of affinity of Gag proteins for gRNA is not a property of Gag in general but specific only for F-Gag. If U-Gag would have increased binding affinity for gRNA (due to allosteric coupling) and an increased ability to associate with other Gag molecules (due to exposed CA domains) then the F-to-U transformation could act as a regulatory assembly gateway that prevents or retards capsid assembly. In this paper, we report on MD simulations of the interactions between individual domains of Gag. Based on the earlier simulation study [28], we first hypothesized that the physical mechanism that stabilizes the F-Gag state involves attractive interactions between the MA and CA-CTD domains that should be strong enough to disrupt the hydrophobic Gag-Gag interactions that stabilize hexagonal Gag arrays [28]. Next, the CA domain has exposed hydrophobic residues on its outer surface and homo-dimeric hydrophobic interactions between adjacent, charge-neutral CA-CTD domains are known to make an important contribution to the stabilization of assembled hexagonal arrays of Gag proteins. For this reason, we also hypothesized that the hydrophobic residues of CA-CTD that are assembly-

Figure 2.2.1: MA-CA bound state. Examples of electrostatic interactions are highlighted: (a) the positively charged (blue) R-43 residue of MA (green) is in proximity of the two negatively charged (red) E-344 and E-345 groups of CA-CTD (blue). (b): the (gold) Q59 and Q63 polar residues of MA are in proximity of the (gold) Q-351 polar residue of CA-CTD.

critical also stabilize the MA-CA bound state since that would be an effective way to prevent assembly of Gag proteins. The simulations discussed below on MA/CA-CTD interaction and on CA-CTD dimeric interaction support the first but not the second hypothesis.

2.2 All-atom simulation and Umbrella Sampling

2.2.1 MA-CA bound state

In order to search for an MA/CA-CTD bound state, we constructed a list of possible MA/CA-CTD pairing configurations arranged in terms of binding scores computed using standard protein-protein docking software (see Methods Section). Next, each candidate configuration underwent a short MD simulation. Only one of the configurations, the one shown in Fig.2.2.1, survived the test. Electrostatic interactions dominate the binding interface of this configuration. It included the polar CA-CTD residues R294, V297, M347, T348, Q351 and the two negatively charged E-344 and E-345 residues of CA-CTD as well as the polar MA residues F44, Q59, Q63, P66, S67, T70 and the positively charged residue R43. Figure 2a highlights the charged residues and Fig.2b some of the polar residues. The polar MA residues of the binding interface are preceded by a long sequence of positively charged residues (Lys-15, Lys-18, Arg-20, Arg-22, Lys-26, Lys-27, Lys-32, Lys-39, and Arg-43) while it is followed by two more (Lys-95, and Lys-98). The free energy difference between F-Gag and U-Gag of course could also involve the other domains of Gag. To test whether this MA/CA-CTD bound state survives as part of a complete Gag protein, we carried out simulations of full Gag with an initial state in which the MA/CA-CTD pair is bound (simulations were limited to 40 ns durations because of the very large size of the Gag protein). The bound state

remained intact over the duration of these runs.

To quantify the strength of the bound state, we computed the Potential of Mean Force (PMF), using the umbrella sampling method. The results are shown shown in Fig.2.2.2. The horizontal axis (X) is the MA/CA-CTD center of mass (COM) separation. The PMF

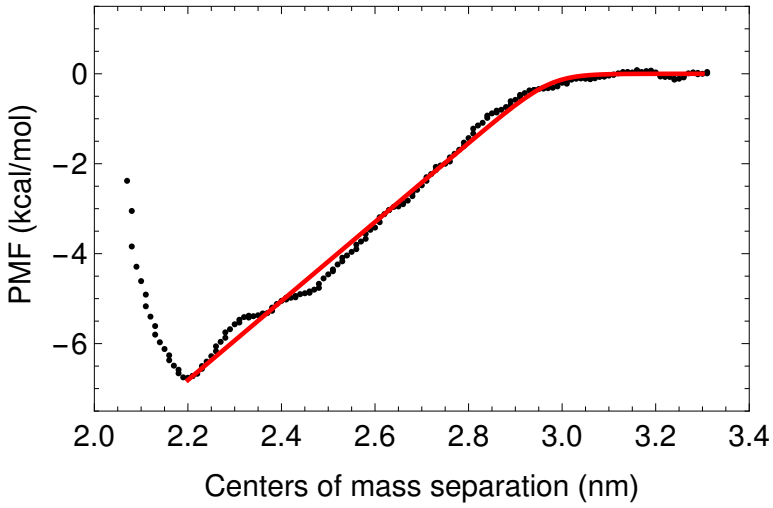


Figure 2.2.2: Left: Potential of Mean Force $V(X)$ between the MA and CA-CTD subdomains (Gold). Vertical axis: PMF in kcal/mole. The horizontal axis X axis is the distance between the centers of mass of the MA and CA subdomains in nanometers (nm). Red line: fit to $V(X) = V_0 - k_B T \ln\left(\exp(\beta\Delta U - F(X - X_0)) + 1\right)$.

has a potential well with a depth ΔV_{MC} of about 7 kcal/mole ($\simeq$ 12 k_BT) while the COM separation X_0 at the potential minimum is about 2.2 nm. Between 2.2 and 3.2 nm, the PMF increases approximately linearly, after which it levels off. Polar interactions are the main contributers to the attraction in the linear regime of the PMF. The linear part of the PMF translates into to a constant attractive force F of about 44 pN. If the bound state is treated as a two-state mechano-chemical system subject to a constant force F_i then corresponding PMF would have the form $V(X) = V_0 - k_B T \ln\left(\exp(\beta\Delta V - F_i(X - X_0)) + 1\right)$. [1].

[1]The actual ΔV is expected to be somewhat less than the computed value because the sampling umbrella potential used in the computation of the PMF has spherical symmetry and thus constrains the motion to the CA-CTD COM in all three directions, not just along the COM separation vector. For COM separations larger than 3.3 nm, the actual ΔV should be reduced by the entropic free energy associated with the release of two translational degrees of freedom, which equals $k_B T$ (black dashed line in Fig.2.2.2).

When the computation of the MA/CA-CTD PMF was repeated on approach, with an initial COM separation X increased to 3.2 nm but with the same relative orientation of the MA and CA-CTD domains as in the bound state, then the simulations failed to equilibrate. This was found to be due to rotational Brownian motion causing orientational misalignment between the MA and CA-NTD domains. The MA/CA-CTD interaction potential is apparently highly directional: the two domains need to be lined up precisely in order for a bond to form. The interaction appears to have a *ratchet-like* character with kinetic traps appearing on approach but not on separation.

To gain insight into the effect of the COM displacement on the bonding between inidividual residues, we made histograms of the separation between the centers of mass of the positively charged R-43 residue of MA and the negatively charged E-344 residu of CA for different values of the mean COM separation $< X >$ of the two domains. The COM of the MA domain was fixed while the COM of the CA domain was subject to a parabolic umbrella potential. For $< X >= 2.3$ nm, which is at the bottom of the PMF, the histogram were fitted by two Gaussian with maxima at 0.79 nm, respectively, 0.83 nm. The width of the Gaussians was comparable to the width of the COM of the CA-CTD domain in the umbrella potential. When the COM separation was increased to 2.7 nm the histogram retained the same general shape but displaced by about 0.4 to 0.5 nm while it also is stretched. The dispalcement is comparable to the increase of the COM separation. The stretching is attributed to a combination of a rotation of the bonding direction and a structural deformation. The histogram were fitted with two Gaussians centered at 1.20 nm and 1.34 nm. The electrostatic bond apparently was still intact though the mean separation between the two residues had increased. Finally, when the COM separation was increased to 3.2 nm, a single peak in the histogram was observed at 3.1 nm, which is considerably in excess of the COM displacement. The natural interpretation is that the electrostatic bond had snapped. The extended tail toward states with smaller separations indicates that there still were fluctuations towards states with residual electrostatic bonding.

2.2.2 Homodimeric CA-CTD Interactions

In order to place the intermolecular MA/CA-CTD bond in context, we compared it with the intermolecular bond between two Gag proteins that are part of the Gag lattice of the immature virion. The sub-tomographic study of ref.[28] indicates that the assembly-critical residues W316 and M317 of one CA-CTD domain form a homo-dimeric hydrophobic contact with the W316 and M317 domains of the CA-CTD domain of an adjacent Gag protein. We computed the PMF of two CA-CTD domains with the starting state of the simulation produced by excising a pair of CA-CTD domains from across the two-fold symmetry sites of the hexagonal lattice. The pair first was thermally equilibrated by an 80 ns MD simulation (see Fig.5), during which the bond remained intact, followed by a measurement of the PMF. The PMF of the CA dimer has a potential well with a depth ΔV_{CC} of about 1.8 kcal/mole (or about 4 k_BTT) and a range of about 0.6 nm. Unlike the MA/CA-CTD PMF, there is no extended region with constant slope. The CA-CA binding free energy ΔV_{CC} is about three times smaller than the MA/CA-CTD binding free energy. The MA-CA electrostatic interaction is apparently sufficiently strong enough to disrupt the hydrophobic bonds that stabilize the hexagonal Gag lattice of the capsid.

2.2.3 Conclusion

The simulations support the proposal that the Gag protein has a bound state (F-Gag) with a binding energy that exceeds that of the hydrophobic interactions that stabilize the hexagonal Gag lattice. Against expectations, the simulations also indicate that this bound state has an electrostatic character. It already is well appreciated that electrostatic interactions play an important role for the state of the Gag protein in general and for RNA selection in particular [32]. MA has a non-specific affinity for RNA molecules and the plasma membrane believed to be due, at least in part, to the large number of positively charged residues of MA. Separately, the NC domain appears to bind to RNA by a combination of non-specific electrostatic interactions that involve positively charged residues of NC and specific hydrophobic interactions [32]. These electrostatic interactions involve charged residues but

the MA/CA-CTD binding interface is largely charge neutral. The MA/CA-CTD bound state we are reporting on is mostly stabilized by dipolar interactions. Because binding between dipolar molecules is weaker than that between molecules with opposite monopole charges the total MA/CA-CTD binding energy of 7.2 kcal/mole necessitates multiple dipolar contacts, which we indeed found to be the case. Because the MA/CA-CTD bond is held together by a significant number of weak polar contacts that collectively determine the binding free energy, we should expect that the extended polar MA/CA binding interface is conserved in terms of mutations. There indeed is a stretch of 14 residues known as the *major homology region* (MHR) that appears to be highly conserved, not only for HIV-1 Gag variants, but also across different retroviruses. Deletion or mutation of MHR residues leads to less virion production and higher levels of Gag oligomerization on the membrane. The MHR corresponds to the residue numbers 285-304 in Fig.1 so it does include the CTD-CA residues that are involved with the formation of the bond.

The simulations disproved our second hypothesis concerning a key role for the assembly-critical hydrophobic residues of the CA-CTD domain. The hydrophobic residues of CA-CTD actually remain fully exposed to the aqueous environment. This has the interesting consequence that two F-Gag proteins should be able to dimerize through hydrophobic interactions and possible form VLPs. The F-Gag bound state is sterically quite different from the conformation of Gag proteins that are part of a hexagonal lattice so aggregates of F-Gag should be quite different from HIV-1 virions or the virion-like VLPs. In-vitro self-assembly studies that involve mixing RNA and Gag in physiological salt buffer, report formation of $\simeq$ 30 nm diameter particles in addition to the $\simeq$ 130 nm diameter virion-like VLPs [Campbell, 1999]. We propose that these small-sized VLPs, which are relatively unstable, are composed of F-Gags linked by the exposed CA-CTD hydrophobic residues.

According to the physics of aqueous electrostatics, binding between oppositely charged macroions should be weakened by the addition of polyvalent ions while addition of monovalent salt should have the same effect. The addition of negatively charged tRNA [33] and of IP6 (cyclic penta-phosphate) groups [34] both increase the binding specificity of Gag for

gRNA. Separately, the addition of the IP6 molecule stabilizes normal-sized VLPS over the small-sized VLPs. A natural interpretation would be that the association of the negative polyions with the positive MA residues weakens the bound state but this is not so obvious because of the presence of the large number of MA positive residues that lie outside the binding interface: binding of IP6 or tRNA to MA would be expected to involve mainly these positive MA residues. However, some of the positive MA residues are immediately adjacent to the polar residues of the binding interface so the MA/CA-CTD bound state still could be disrupted by short-range allosteric coupling between the charged and polar residues when tRNA or IP6 binds to the positively charged MA residues. Related to this, if the MA domain would be more flexible in the unfolded state than in the bound state, then this also could favor disruption of the bound state because that flexibility would facillitate the binding of the tRNA and IP6 groups to the positively charged residues. These possibilities can all be investigated by MD simulations and we plan to do this. Finally, long-range allosteric interactions between the NC and MA/CA parts of Gag also could play a role in terms of the interaction of MA with tRNA and IP6. Checking this requires more complex simulations of the whole Gag proteins.

The effect of increased monovalent salt concentration on the RNA binding specificity of Gag deserves additional discussion. The authors of Ref.[32] report that increased salt concentration increases binding specificity. They interpret their results in terms of the dependence on salt concentration of the binding of the NC sequence to the ψ sequence, an important packaging signal of gRNA. They argue that at lower salt concentration generic electrostatic interactions of RNA with the MA and NC groups collectively produce a folded state. Increased salt concentration weakens this state in favor of specific hydrophobic interactions of NC with the ψ sequence. The MA domain can interact with the plasma membrane in that case. The authors of a second study [33] confirm that increased salt concentration enhances binding specificity and also confirm that the competition between non-specific electrostatic interactions with non-specific hydrophobic interactions play an important role. In addition, they report a surprising dependence on salt concentration of the dissociation con-

stant of the binding of Gag to RNA: it is relatively independent of salt concentration at low salt concentrations until it sharply rises at higher salt concentration beyond a threshold concentration. In contrast, one would expect on the basis of aqueous electrostatics that increased salt concentration should immediately weaken the generic electrostatic interactions and that only when non-electrostatic interactions dominate should the dissociation constant be independent of salt concnetration. The authors of [33] interpret this results in terms of a salt-induced conformational change of Gag and propose the presence of a folded state of Gag stabilized by electrostatic interactions. Our simulations would be consistent with this proposal but it is clear that MD simulations of the MA/CA bound state have to be repeated at different salt concentrations and compared with simulations of the NC-RNA interaction before drawing conclusions.

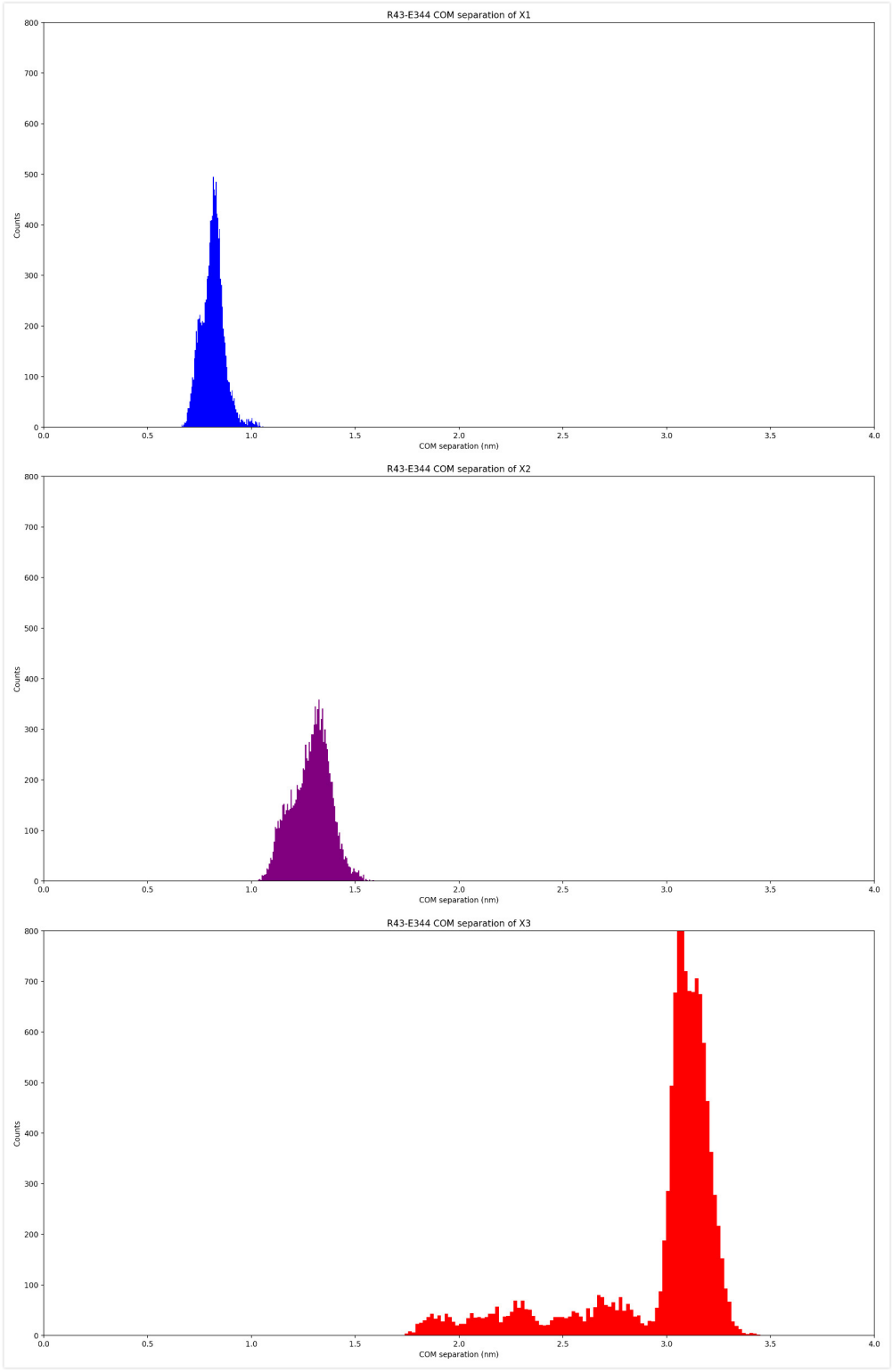


Figure 2.2.3: Histogram of the COM separations of the R-43 residue of MA and the E-344 residue of C for different values of the average COM separation $< X >$ of the two domains. Top: 2.3 nm. Middle: 2.7 nm. Bottom: 3.2 nm

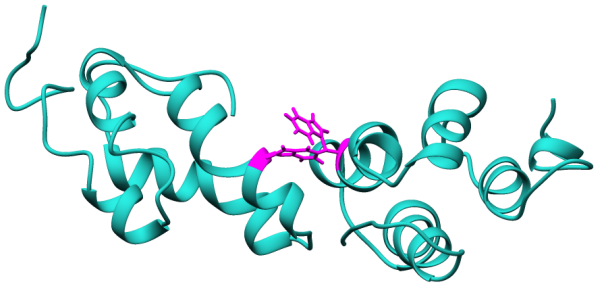


Figure 2.2.4: CA-CA contact after a 80ns simulation The hydrophobic residues W316 and M317 are highlighted.

CHAPTER 3

Diffusion-limited reactions between confined proteins.

3.1 Introduction

In the previous section we studied bonding between the MA and CA sub-domains of the Gag protein where the initial state already was bonded. Using all-atom MD simulations, we obtained an effective potential of mean force. However, the time-scale for the bonding between the MA and CA sub-domains of an isolated Gag protein starting from an *extended* conformation is much larger than what can be probed by MD simulations. The two sub-domains are connected by a long, flexible linker. Kinetics is diffusive and, because the MA-CA boundstate is highly directional, presumably involve many contact events that do not lead to bonding. This clearly is a problem that should be studied at a coarse-grained, mesoscopic level.

The problem of the MA and CA groups locating each other is a special case of a diffusion-limited process in molecular biology. These are quite common for reactions involving enzymes and proteins where a ligand binds to a receptor on a molecule [35]. For the case of reactions on cell membranes, a target molecule may have to diffuse in two dimensions along the membrane until it hits a patch before a reaction can take place. A repressor protein may have to diffuse in one dimension along a DNA molecule until it finds an operator sequence to which it can bind. Alternatively two proteins in contact may have to adjust by rotational diffusion to a conformation with the correct binding pockets lined up [36]; This group of problems often can be simplified in the form of a toy model in which particles diffuse along some boundary. Such models were first proposed nearly a century ago by Smoluchowski [37] and have since been studied extensively in different conditions and aspects to mimic different backgrounds.

In the pioneering work of Berg and Purcell [36], it was assumed that there were small non-overlapping patches distributed over the surface of a sphere. Next, in the Collins and Kimball model [38], a surface was discussed that was partially covered by reactive material. The initial condition of the reactants may be variable. For instance, the small particle maybe initialized at infinity [36] or it directly starts on the surface [39]; it could also be confined within the targeted surface or restricted by a reflective boundary/harmonic potential outside the surface which animate the escaping of a molecule from a cell or compartment [40][41]. Most papers assumed the reaction surface to be a sphere. To complement that, other geometries were also investigated such as a sphere with cavity [42] or the surface of a cylinder et.al [43]. Other extensions are cases where the reactants are rotating with several reactive patches [44].

Almost all studies we just mentioned assumed that the patches are purely reactive while the remaining surface is reflective. Typically, the achievement be made with these conditions is the rate of reaction or *mean first passage time (MFPT)/first hitting time (FHT)* of the process. Complicated surface mechanism with this assumption have been studied. The MFPT of events triggered by arrival of multiple particles instead of a single has also been studied. This can be used to explain the process of vesicular transmitter release in neurons is initiated after the arrival of five calcium ions onto the vesicle sensor et.al [45].

The reactivity of patches is another feature that has been discussed. Variable reactivity can be modeled by introducing and activation energy barrier that needs to be overcome before binding or activity/accessibility of the target. Reactions may happen with a certain probability conditional on two object touching each other. This extension of the conventional diffusion & first hitting time model is named the *Diffusion Limited Reaction Model* [46]. Beyond that, motivated by systems with targets that are not always reactive, stochastic gated models have been developed. In this case, the reactivity is not a constant but may decay or be random. This appears to reproduce the binding site under conformation transitions or catalysis fooling et.al [47]. Finally, the remaining non-reactive surface may not be purely reflective, for example because of generic adhesive interactions. This mechanism could play a big role to shift the kinetics from bulk to surface transport or to an intermediate form

known as *Bulk-mediated Surface diffusion*.

A simple example of this form of transport is the *Lévy walk*. This happens when the hopping of the particle along the surface depends on an adsorption/desorption rate while the particle may resume its surface diffusion with a certain probability. It was shown that under these conditions the particle performs a Lévy walk [48]. Thorough studies with different desorption rates have also been carried out with different patch sizes [49]. This last mechanism, separating a liquid bulk phase and a surface is important in biological process. The exchange between surface&bulk diffusion confers a flexibility to bio-system for a target search process.

An important mathematical result in this area should be mentioned here. Grebenkov [50] has proposed a new formalism to describe the bulk-surface diffusion process with mixed boundary conditions. As we just reviewed, conventional description relies on bulk diffusion D and reactivity κ on the surface, where the surface reaction mechanism are incorporated via boundary condition, i.e. mechanism on patches. However, in the new paradigm he proposed, instead of using a single Langevin equation, he used a stochastic equation:

$$dX_t = \sqrt{2D}dW_t + n(X_t)dl_t \tag{3.1.1}$$

to describe two process simultaneously. In equation 3.1.1, the X_t is the position of particle, $n(X_t)$ is a boundary normal vector and importantly l_t is boundary local time, it only increases when the particle touches the boundary (reactive region). The intriguing property of the new equation is that it separately describes the bulk diffusion X_t and surface encounter l_t, The full propagator P_0 is the joint probability of the two processes. The two random processes can have their own stopping time or first hitting time. For example, the particle may die spontaneously as a Poisson process, i.e. the particle will die with the same probability in each time spot. The boundary process can be set in same manner, and the reaction rate κ it self can be turn to a variable, leading the stopping time for the local boundary to be an integral of the distribution:

$$\Phi(l) = e^{-\int_0^l dl' k(l')} \tag{3.1.2}$$

Such alternative description allows one to model complicated surface conditions. For instance, for the case that the patch reactivity is not a constant, one can set it as a distribution to including this feature in one analytic description.

After this short review of bulk-surface diffusion and reaction problems, we return to the MA-CA problem and develop our own approach.

3.2 Basic settings of the coarse-grained model

We studied a model where a sphere and a moving point particle represented the two subdomains of the protein. A sketch is shown below. The harmonic potential between them models the confinement by the tether models the polymer chain between MA and CA subdomains. Similar coarse-grained models could be used to address the broader question of reactions between weakly-interacting proteins at high concentrations or, essentially equivalently, reactions between a few weakly-interacting proteins confined inside a small volume. Next, we also included a Lennard-Jones (LJ) potential between the two objects. The attractive part represents non-specific affinity, for example due to van der Waals attraction. The repulsive part of L-J is the hardcore of the sphere.

The sphere has two portions, one is reflective, the other is reactive, which is our simplified binding site. In the simplest case, the southern hemisphere is the reactive part while the northern hemisphere is the reflective part. For the reactive part, if the particle reached a certain point inside the well of the L-J potential then the simulation was stopped.

Below, we first offer a theory for the steady state distribution of the system. Then we extend this theory by implement Brownian dynamics. The detailed setting can be found in section 3.4.

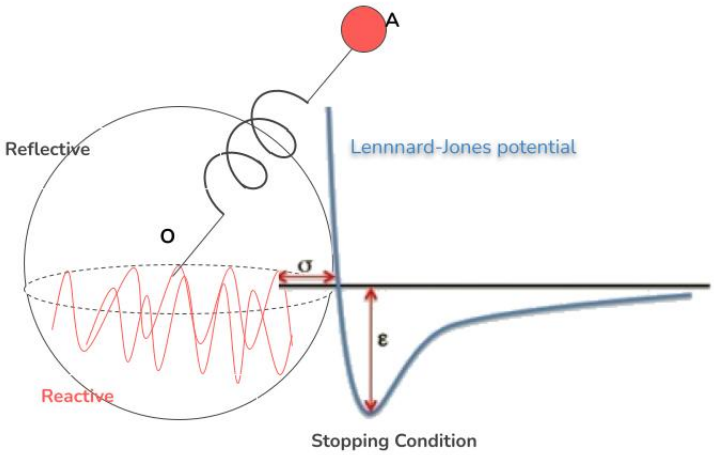


Figure 3.2.1: Toy representation of the coarse grained model

3.3 Fokker-Planck equation and steady state distribution

We've used two analytic ways to derive the steady state concentration profile. The first one was turning the stochastic Langevin equation to a deterministic Fokker-Planck equation to find the probability distribution of particles. For the second way, the basic assumption was that there always was one particle. Whenever the simulation was stopped, a new particle was introduced at the maximum of the probability distribution. Then, by considering the conservation of current at equilibrium distance i.e. 20 in the case of $k = 1/400$, we could find an expression for probability distribution of particles. The resulted expression of the distribution is:

$$P_1(r) \propto e^{-\beta\phi(r)} \int_{r_0}^{r} \frac{e^{\beta\phi(r')}}{r'^2} dr'$$

The plotted curve is shown below:

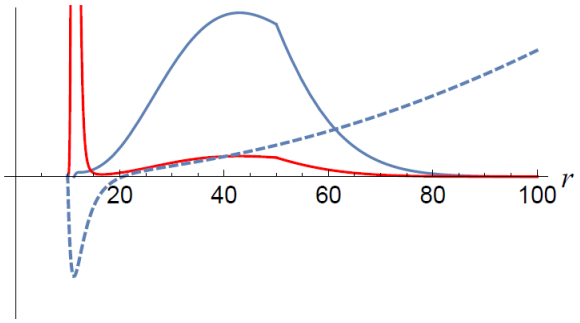


Figure 3.3.1: The blue dashed line is the L-J potential. The blue and red solid curves are steady state distributions with absorbing surface at r = 10 (σ) and r = 11.2 (min of L-J) respectively

3.4 Numerical model of a diffusing particle

3.4.1 Diffusion process

The dynamics of a mesoscopic system in continuous space can be described by *Diffusion process*. In solution system, the diffusion process most of time related to Brownian motion, which were investigated by Langevin, Einstein [51] and Smoluchowski at the beginning of last century [52]. Then later proved experimentally by Perrin [53].

In diffusion process, the effect from surrounding medium like the collision by solvent molecules are integrated together as a thermal noise. Mathematically, this process can be described by a *Langevin equation*:

$$m\ddot{\vec{x_t}} = -\nabla U(\vec{x_t}) - \gamma m\dot{x}_t + \sqrt{2m\gamma k_B T}R(t) \tag{3.4.1}$$

where $U(\vec{x_t})$ is the potential energy of the system, usually in equilibrium condition it is time invariant. γ is the friction coefficient and $R(t)$ is the white noise, which should be a stationary Gaussian process and satisfying $< R(t)R(t') >= \delta(t-t')$. Noticed the noise term represents the thermal fluctuation of the system at a certain temperature T, thus it directly samples canonical ensemble.

The overdamped case of Lagevine dynamics is *Brownian dynamics*, in which the accel-

eration was ignored: $\gamma m\dot{x}_t = -\nabla U(\vec{x_t}) + \sqrt{2m\gamma k_B T}R(t)$. Below, we will use a discretized Brownian dynamics with dimensionless units to implement the numerical simulation.

3.4.2 Basic settings

We carried out a Brownian dynamics to extend our analytical theory. The dynamics was governed by following updating rule:

$$X_{n+1} = X_n + V_n \Delta t \tag{3.4.2}$$

$$V_{n+1} = V_n(1 - \Delta t) + F_n \Delta t + \Delta n \tag{3.4.3}$$

the X_n, V_n and F_n are dimensionless displacement, velocity and external force respectively while Δn is dimensionless Wiener noise with a correlation function that is a delta function in time. We leave all detailed derivations to the Appendix section.

To implement the simulation, we first obtain the equilibrium distribution of the system by running Langevin dynamics for a 10,000 steps during which there were no reaction. The distribution without L-J attraction has a wider range than that with L-J as shown below:

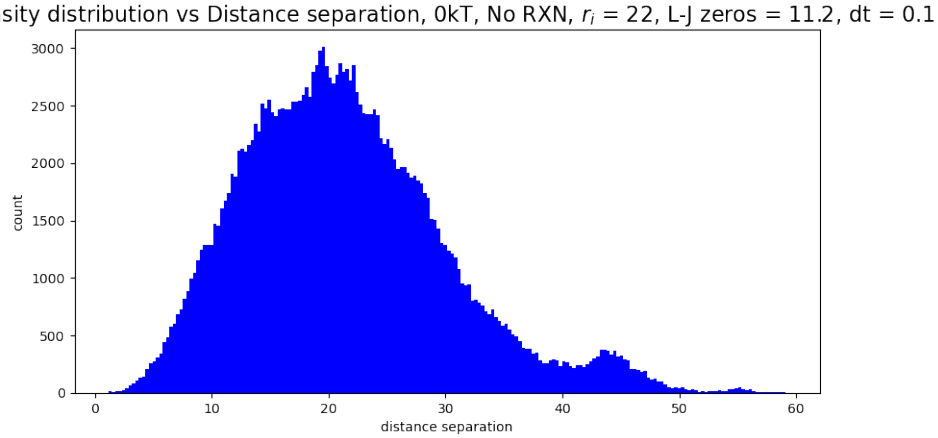


Figure 3.4.1: Equilibrium distribution with pure stochastic process, without stopping condition

The most probable distance for this equilibrium distribution was used as the starting point or source of the *first hitting model*. New particles were generated after a particle

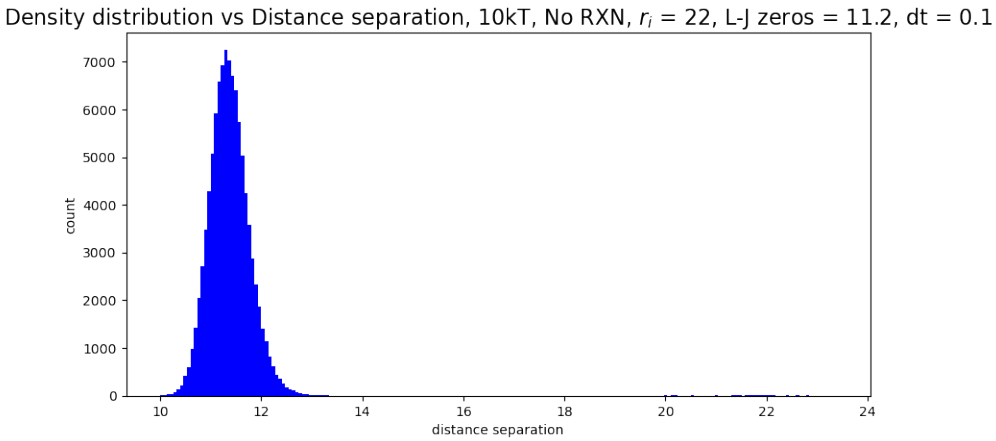


Figure 3.4.2: Equilibrium distribution with 10 kT L-J, without stopping condition

reached the stopping location of the inside the L-J potential of the reactive part of the sphere. For the case that the whole surface was reactive, the resulting distribution had a good agreement with analytic solution as shown below:

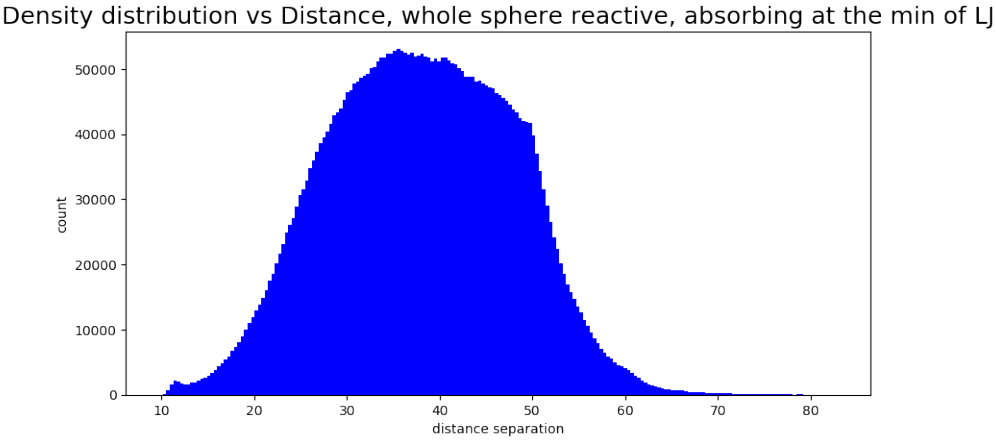


Figure 3.4.3: Steady state distribution with whole sphere reactive, stopping condition set at min of L-J

The distribution of absorbing surface at r = 10 also agree with analytic solution with a significant peak near the boundary (the red curve in figure 3.3.1):

Noticed that above plots are both under the condition of perfect reactive sphere (whole sphere are reactive with probability 1). Thus, if we set the stopping condition at the min of L-J, but with fraction of surface reactive with a probability, the distribution tends to act

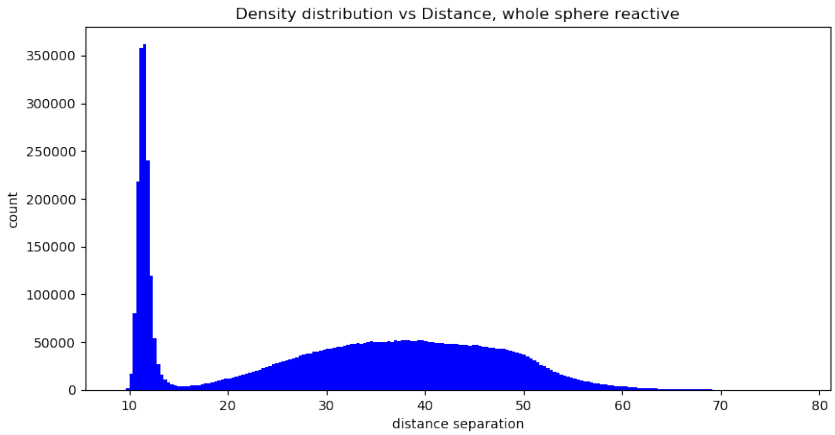


Figure 3.4.4: Steady state distribution with whole sphere reactive, stopping condition set at $\sigma = 0$

like the second plot, in which the reactive probability mimics the energy barrier.

3.5 Pure diffusive and incoming current with an edge field

We initialized the behavior of particles on the surface for the case that it diffuses in from infinity. The fixed spheres as one reactant is divided into two portions (3.5.1), one is perfect reactive and the other is purely reflective.

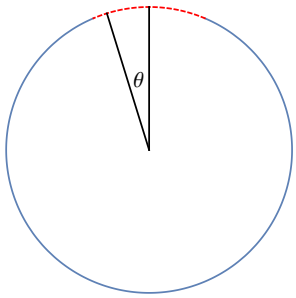


Figure 3.5.1: Two portions of the circle

By utilizing Fick's law and Neumann Boundary Condition, we can find the surface density of particles. In which the current density is of the form $\vec{j}(\vec{r}) = -c\vec{\nabla}n(\vec{r})$, where $n(\vec{r})$ is the density of the source particles that is being delivered. The boundary conditions at the surface

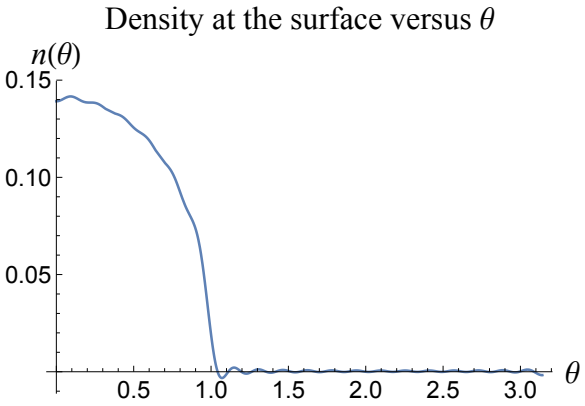


Figure 3.5.2: Surface density with a reflective boundary

give rise to the following restrictions on the density:

$$n(1,\theta) = 0 \qquad 0 < \theta < \theta_0 \tag{3.5.1}$$

$$\left.\frac{\partial n(r,\theta)}{\partial r}\right|_{r=1} = 0 \qquad \theta_0 < \theta < \pi \tag{3.5.2}$$

Furthermore, because of the symmetry of the system, we know that we can express $n(\vec{r})$ in the following form outside of the sphere:

$$n(r,\theta,\phi) \equiv n(r,\theta) = n_0 + \sum_{l=0}^{\infty} \frac{a_l}{r^{l+1}} Y_l^0(\theta,0) \tag{3.5.3}$$

By expanding the spherical harmonic by Legendre polynomials, $P_n(\cos\theta)$, via

$$Y_l^0(\theta,\phi) = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2l+1}{2}} P_l(\cos\theta) \tag{3.5.4}$$

The complete set of functions fulfill the boundary conditions in the two regions by its orthonormal property.

By truncating the equation with a complete set of 80 Legendre polynomials, one can solve the equation analytically and get the results of density and incoming current on the surface, as shown in 3.5.2 and 3.5.3:

It's clear that the density is zero in the region in which the sphere is absorbing, agree with our boundary condition. The incoming current also have a good agreement, which is zero

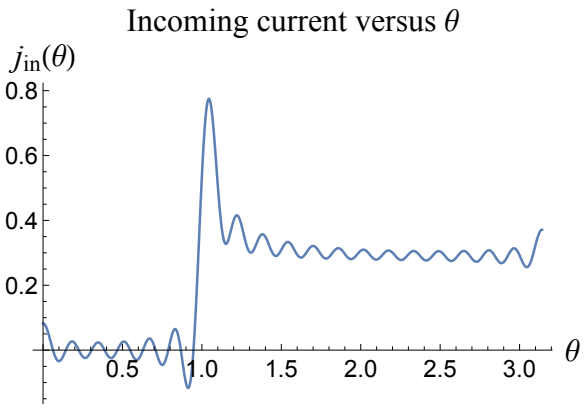


Figure 3.5.3: Surface density with a reflective boundary

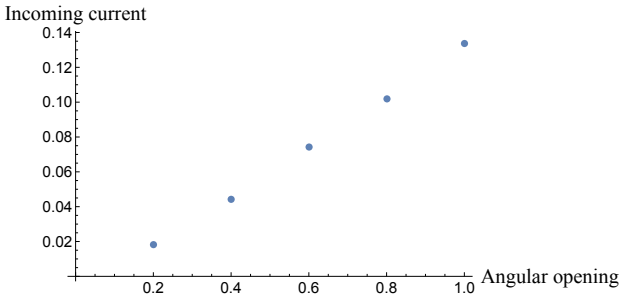


Figure 3.5.4: Reaction rates respect to different open angles by calculating incoming current

in the region in which the sphere is reflecting (all particles tends to flows into the absorbing region due to concentration difference).

Finally, one can set different portion of the sphere to investigate the relation between incoming current and open angle. The result shown in 3.5.4 nearly a straight curve, but as log-log plot fitting shows, the exponent is 1.24. **One should not expect the naive linear proportionality of Angle-Rate relation with large angles, which breaking the assumption of Berg-Purcell model**.

The density and incoming current at the vicinity of the boundary between absorbing and purely reflective region can be also described by Laplace's equation by *conformal mapping* [54].

By investigating an analytic function of the complex variable $x + iy$, where x and y are

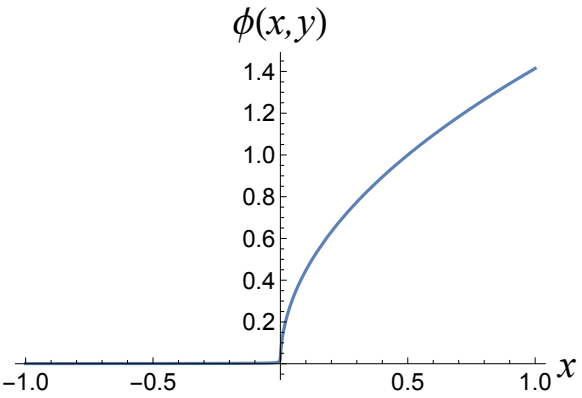


Figure 3.5.5: Complex function $\Phi(x, y)$ to represent surface density

real that has the proper behavior along a line of singularities. The function is actually simple as:

$$\sqrt{x+iy} = \frac{1}{\sqrt{2}}\sqrt{\sqrt{x^2+y^2}+x} + \frac{i}{\sqrt{2}}\sqrt{\sqrt{x^2+y^2}-x} \tag{3.5.5}$$

As the function above is a harmonic function, it's real and imaginary parts are both solution of Laplace's equation:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) \frac{1}{\sqrt{2}}\sqrt{\sqrt{x^2+y^2}+x} = 0 \tag{3.5.6}$$

When $y = 0$, the behavior of real part of (3.5.5) is interesting. When $x > 0$, the argument of the outer square root is $2x$. When $x < 0$, the argument is zero. In fact, we have a function that, along the real axis is zero when x is negative and is proportional to $\sqrt{x}$ when x is positive, which is similar to the behavior of density at two sides of the boundary. Making y non-zero smooths out this behavior. The figure 3.5.5 is a plot of the real part (without the square root of 2) for $y = 0.001$ and constant.

We expect the derivative of the potential function with respect to y have similar behavior as incoming current. The derivative is:

$$E_y(x, y) = \frac{y}{2\sqrt{x^2+y^2}\sqrt{\sqrt{x^2+y^2}+x}} \tag{3.5.7}$$

Taking the limit $y \to 0$, we find that for $x > 0$ this function is zero, while for $x < 0$ is is proportional to $1/\sqrt{x}$, which similar to the incoming current we presented above as expected. Again, the function smooths out when y is non-vanishing as the figure shown in figure 3.5.6:

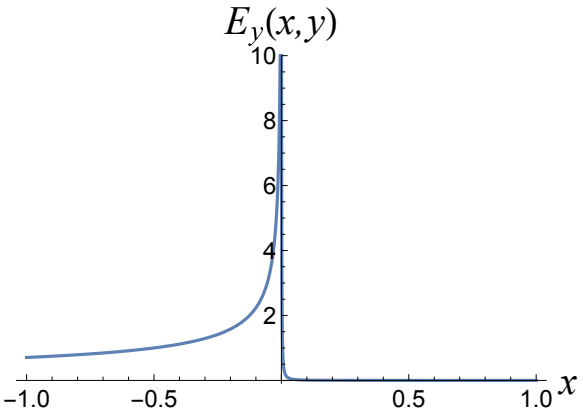


Figure 3.5.6: Differentiation of complex function Φ (x,y) to represent incoming current

3.6 Analytical model of surface hopping and evaporation

After simply assumed the particles are purely reflective on the surface of the sphere, we now structured a model which allows particles can be absorbed and evaporated with a certain parameters, which plays a role as Lennard-Jones potential allowing the particles to move in a finite range above the surface. Notice that the film will have variable thickness, but we further assume that the structure of the film does not affect the diffusion current density of the incoming particles, i.e. tangential moving along the surface.

We have the following equation governing the time dependence of $c(x,t)$:

$$\frac{\partial c(x,t)}{\partial t} = S - \alpha(x)c(x,t) + D\frac{\partial^2 c(x,t)}{\partial x^2} - \epsilon c(x,t) \tag{3.6.1}$$

Here, S is the constant incoming current density of particles, while $\alpha(x)$ is a measure of the permeability of the substrate and ϵ is the rate of evaporation. We will assume that the film extends from $x = 0$ to $x = L$, and that the film is uniformly permeable from $x = 0$ to $x = L/2$. We will impose Neumann boundary conditions on the density at the extrema of the film, which means we can express $c(x,t)$ as follows.

$$c(x,t) = \sum_{n=0}^{\infty} C_n(t)\cos(n\pi x/L) \tag{3.6.2}$$

We could analyze the steady state situation, which is described by Eq. (3.6.1) with the right hand side equal to zero. We start by integrating the left hand side of the equation from

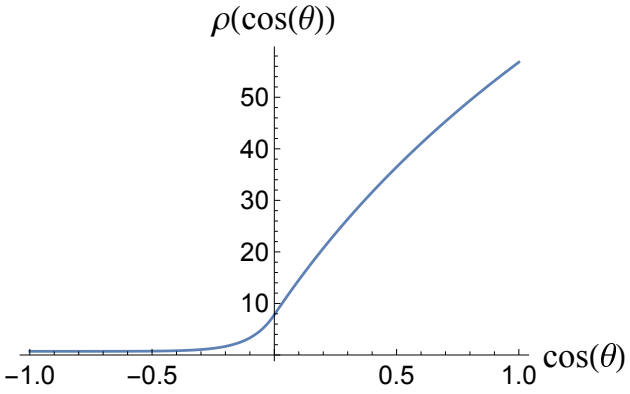


Figure 3.6.1: The surface density profile calculated by equation 3.6.1, by turning off evaporation rate

$x = 0$ to $x = L$, where the evaporation will be turned on later.

$$\begin{aligned} 0 &= SL - \int_0^{L/2} \alpha c(x)\, dx \\ &= SL - \sum_{n=0}^{\infty} \int_0^{L/2} \alpha C_n \cos(n\pi x/L) dx \\ &= SL - \alpha C_0 L/2 - \alpha \sum_{n=1}^{\infty} C_n \sin(n\pi/2) \frac{L}{n\pi} \end{aligned} \tag{3.6.3}$$

(See more details in appendix)

The plots 3.6.1 show the results of solving the equations by truncation with n going up to 199.

The diffusion current on the surface can be calculated by taking derivatives of the density file. The resulted curve is shown below, where the current at the vicinity of reactive/reflective boundary is maximized, consistent to our intuition. The surface hopping mechanism is analyzed below by the Markov State Model.

The above calculation was carried for assuming a flat surface. To extend the model to the surface of a sphere, we used spherical harmonics. Taking consideration of the symmetry of the system, we have spherical harmonic: $Y_m^l(\theta, \phi) \propto P_l(\cos\theta)$. In fact, the proper basis set consists of the normalized Lagrange polynomials

$$f_l(\theta) = \sqrt{\frac{2l+1}{2}} P_l(\cos\theta) \tag{3.6.4}$$

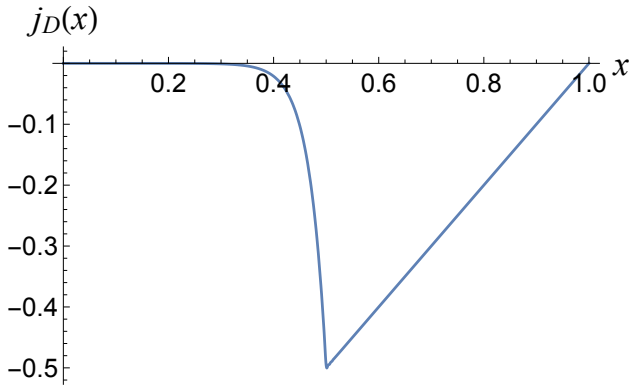


Figure 3.6.2: the incoming current profile calculated by equation 3.6.1, by turning off evaporation rate

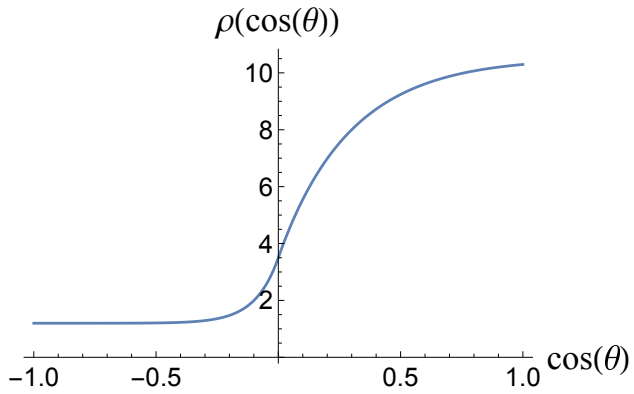


Figure 3.6.3: The surface density profile calculated by equation 16, with finite evaporation rate $\epsilon = 0.13$

If we expand (3.6.1) in this basis set, we can solve for densities and diffusion currents similar to those calculated for the sinusoidal basis set. We replace θ as a variable by $x = \cos\theta$, so that we can write

$$c(x,t) = \sum_l C_l(t) f_l(x) \tag{3.6.5}$$

This expansion gives us the ability to treat this as a non-homogeneous system, which will be useful for the analysis of the eigenvalue structure. The detailed solution of above equation can be found in the appendix.

By truncating the equation above and turning on the evaporation parameter, we could obtain the surface concentration profile.

Finally, we need to take three dimensional diffusion into account. If one assumes that evaporated particles re-condense uniformly, then we can effect that outcome by restoring the value of the x-independent $l = 0$ contribution, which is the only one that integrates to a non-zero value. Thus, the new equation set is B.0.2 (in appendix) for $l \neq 0$ and the equation below for $l = 0$.

$$\begin{aligned} 0 &= S - \alpha \sum_{l'} C_{l'}^{(0)} \int_{-1}^{x_0} f_l(x) f_{l'}(x)\, dx \\ &= S - \alpha \int_{-1}^{x_0} f_0(x) \rho(x)\, dx \\ &= S - \alpha \frac{1}{\sqrt{2}} \int_{-1}^{x_0} \rho(x)\, dx \end{aligned} \tag{3.6.6}$$

Mathematical details are in the appendix.

As just pointed out, when we assume that all evaporation is recovered uniformly over the surface, we end up with a rate of absorption by the sphere that is independent of both the rate of evaporation and the size of the reaction zone. This follows directly from the steady state equation for the $l = 0$ mode of the set (3.6.4), and it is exactly consistent with the constant reactivity we set in the numerical model (see section below).

3.7 Reaction rates and eigenvalue structures from the analytical model

We could take derivatives of the calculated density files to get the incoming current as we did in the section 3.5. The current in steady state is the net flux in the system, that is the reaction rate based on our setting. List of reaction rates with different open angles and evaporation parameters is shown below. Notice that the reaction rate may be smaller with a larger attractive potential (smaller evaporation rate), which means attraction is not always helpful. We will further explain this in detail numerically in the later section.

We then included time-dependence. The function $c(x, t)$ obeys the equation:

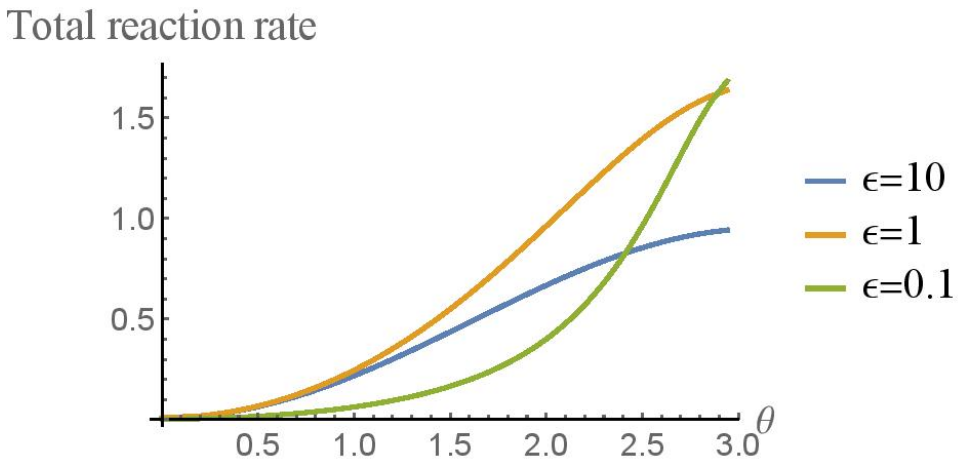


Figure 3.7.1: Reaction rates with respect to different open angles (each curve) and hopping rates, calculated by derivatives of density profile

$$\frac{\partial c(x,t)}{\partial t} = S(t) - \alpha(x)c(x,t) + D\frac{\partial}{\partial x}\left((1-x^2)\frac{\partial c(x,t)}{\partial x}\right) - \epsilon c(x,t) \tag{3.7.1}$$

In this description, both $S(t)$ and $c(x,t)$ are time-dependent on the number of particles in the vapor and film respectively. This relationship between the film and the vapor phase turns the full equation set into a set of homogeneous linear equations. The resulted eigenvalues and eigenvectors correspond to dynamics of the process, i.e. the time dependence of the film concentration. This could be explained as the rate matrix, the eigenvalue zero corresponds to steady state and all the other eigenvalues represent the dynamics before reach it. Figures shown below are the time evolution of the surface concentration and the selected eigenvectors which imply the distrubution changed over time.

Numerically, the concentration change on the film can be directly reproduced by obtaining the time series of the simulation data. The eigenvalue structures also have analogy in rate/transition matrix in Markov State meaning. One can find the resulted eigenvalues and eigenvectors on the next page. We will defer the discussion with more details in the following sections.

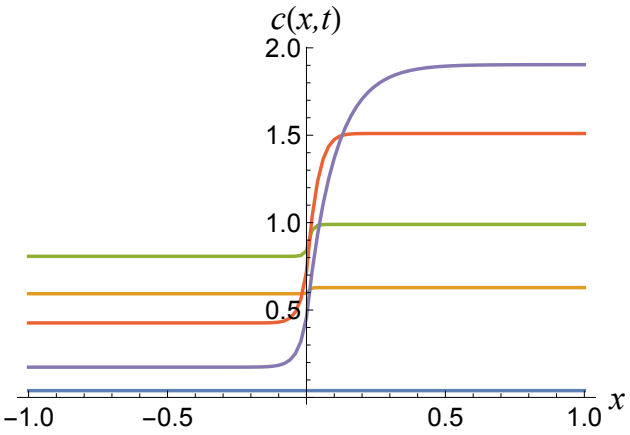


Figure 3.7.2: The time evolution of the density $c(x,t)$ when all particles start off in the vapor phase. On the right-hand side of the plots, the height of the density curve (non-reactive region) increases monotonically with time. On the left-hand side, the density increases at first, then decreases, reflecting the larger rate of loss in the reactive region

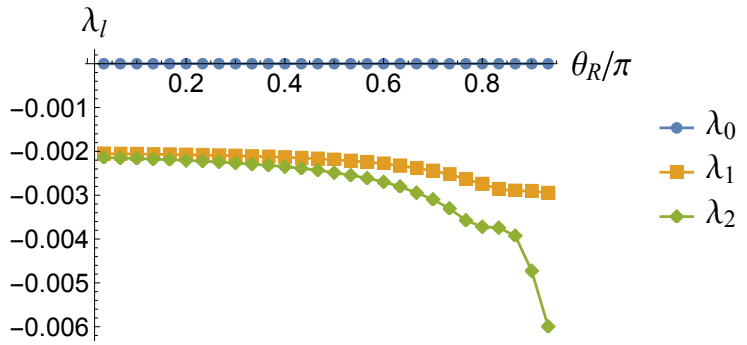


Figure 3.7.3: Three top eigenvalues of the homogeneous function

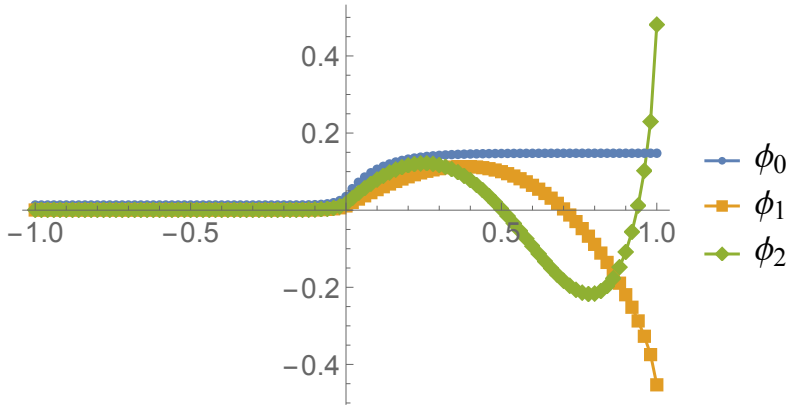


Figure 3.7.4: The eigenvectors associated with the three smallesr eigenvalues.

3.8 Importance sampling and mathematical approaches review

3.8.1 Markov State Model (MSM)

The evolution of many-body physical systems described by Newtonian Mechanics is *deterministic* at an atomistic level in the sense that the state of a system at a certain time is determined by the initial conditions. However, at the level of biologically relevant interacting macromolecules in an aqueous medium an enormous number of degrees of freedom are involved and a deterministic description becomes cumbersome and impractical [55]. The experience has been that the evolution of such systems is best described in terms of a limited set of collective variables subject to thermal noise, describing the effect of, say, the collisions of solvent molecules with the macromolecules. The effect of the noise is very important and probability theory, the theory of stochastic processes, statistical physics and statistical tools all play an important role in the description. The proper choice of the collective variables is an important question in this context. We will call such a description *mesoscopic*, distinguishing it from a microscopic, atomistic descriptions on the one hand and, on the other hand, from macroscopic descriptions such as equilibrium and non-equilibrium thermodynamics or the Navier-Stokes equation [56], when the effects of stochastic fluctuations are less apparent.

The Markov State Model (MSM) is an example of an important statistical tool that can be used to identify the collective variables [57]. Briefly, MSM is a representation of a *Master Equation* that determines the time evolution of the occupation probabilities of a Markov chain [17], as defined below. Before we dive into the implementation of MSM, we first review the MSM in the context of stochastic processes. At typical mesoscopic time scales (e.g., nanoseconds to microseconds). One expects that a mesoscopic system has sufficient time to decorrelate from its initial state because of the effects of the thermal fluctuations of the aqueous solvent that have picosecond correlation times.

The dynamics is coarse-grained in the form of transitions between several distinct conformations, which we will call *macrostates*. In the MSM it is assumed that the number

of such macrostates is finite, says N. Under these assumptions, the simplest form of dynamics is a discrete *random walk* from state to state in which the steps are described by a time-dependent random variable X_t with t an integer representing discrete time. The formal definition of the Markov process is as follows. Suppose the random variable can adopt N possible values (i.e., states) $\{i_j\} \equiv \{i_0, i_1...i_n, ...i_N\}$. Define $P_t(X_t = i_n)$ to be the time-dependent probability that the random variable occupies state i_n at time t. For an MSM, the occupation probabilities at time $t + 1$ are completely determined by the state of the system at time t. For a Markov process the probability of a transition between two states is independent of time. Define the conditional probability that the process is in state j at time $t + 1$ given that it is in state j at time t:

$$P_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) \tag{3.8.1}$$

Gathering all transition probabilities together between all paired states, one can create a transition matrix $\mathbf{Q}$, whose i^{th} row and j^{th} column entry is the conditional probability P_{ij}.

The transition matrix we just defined is the heart of a Markov model. First, it completely governs the evolution of stochastic process. For a given initial distribution of states $\mu_\mathbf{0}$, the distribution after n time steps is $\mu_0\mathbf{Q}^n$, which is known as the Chapman-Kolmogorov equation. Second, the eigenvalue structures (also known as 'spectral decomposition' in some text books) can be analyzed. The following lemma is the mathematical foundation of MSM:

Lemma 3.8.1 *The eigenvalues of Q are bounded by 1, that is* $|\lambda_i| \leq 1$*;*

Lemma 3.8.2 *Q always has the eigenvalue* $\lambda = 1$

Lemma 3.8.3 *Irreducibility of Q is equivalent to the property that every pair of states in the state space communicates with each other. This means that a state of ergodicity can be reached.*

We refer again to [57] for the proofs of these lemmas. Briefly, for lemma 1&2, the (left) eigenvalue 1 corresponds to the steady-state or equilibrium distribution of states. The

remaining eigenvalues describe the relaxation dynamics of the system. The right eigenvectors have a different meaning associated with expectation values.

So far, the time variable took on discrete values. We now want to generalize it to the case where t is a continuous time variable, corresponding to a continuous time Markov chain (CTMC). Formally, assume a continuous trajectory $\{X_t\}$ on a discrete state space: $\{i_0, i_1 ... i_n, ... i_N\}$. We can write down the transition probability by analogy with the discrete case:

$$P_{ij}(\tau) = \mathbb{P}(X_{\tau+s} = j | X_s = i) \tag{3.8.2}$$

The Chapman-Kolmogrov equation takes the form:

$$\mathbf{P}(\tau + s) = \mathbf{P}(\tau)\mathbf{P}(s) \tag{3.8.3}$$

Define the matrix $K = \lim_{h\to 0} \frac{1}{h}(\mathbf{P}(h) - \mathbf{I})$, where h is the time interval between the two jumps. The Chapman-Kolmogrov equation reduces to:

$$\frac{P(\tau + h) - P(h)}{h} = \frac{P(h) - I}{h} P(\tau) \tag{3.8.4}$$

By definition of differentiation and the commutation of $\mathbf{P}(h)$ with $\mathbf{P}(\tau)$ one can solve the equations $\frac{d\mathbf{P}(t)}{dt} = \mathbf{K}\mathbf{P}(t)$ and $\frac{d\mathbf{P}(t)}{dt} = \mathbf{P}(t)\mathbf{K}$, which is so-called *Kolmogrov forward and backward equation*. The solution is:

$$\mathbf{P}(t) = e^{\mathbf{K}(t)}\mathbf{P}(0) \tag{3.8.5}$$

$\mathbf{K}$ is here called the *generator* or *rate matrix* of the Markov chain. The entries K_{ij} represent the rate of switching between state i and j. The transition matrix is the heart of the continuous Markov State Model.

To implement MSM in a practical case, one should first cluster the state space from the raw data of an experiment or numerical simulation to a short list of macro-states. This

can usually be done using clustering techniques, such as *k-means*, or a geometric partition of the microscopic state space, for example by Voronoi tessellation. Then, one can build a transition matrix $\mathbf{Q}$ of the macro-states that were selected. To build such a transition matrix, one starts from a "count matrix" $\mathbf{C}$. An entry of the count matrix $\mathbf{C_{ij}}$ is the number of transitions from state i to j. These counts are of course a function of the lag time τ, which we defined in the CTMC formulation. This a key parameter that controls the performance of MSM and we will give more details later in section 3.7. After selecting an appropriate τ, the transition matrix is then constructed by dividing the the entries of each row by the summation of the row. More details of the technical setting such as how to obtain sufficient data adaptively can be found in [58], [17] and references therein.

The transition matrix provides important information. For instance, one can get the rate matrix by taking logarithm of transition matrix. The net current between two paired states can be obtained. Finally, instead of analyzing the distribution of the macro-states, one can trace the distribution of reactive trajectories by calculating the relative probability of each possible path from entries of transition matrix along the path, which known as *Transition Path Theory*.

3.8.2 Transition Path Theory (TPT)

Transition Path Theory (TPT) is a tool to statistically describe the reactive trajectories [1], aiming to obtain more information from transition events. Traditional ways of analyzing transition rates such as *Transition Stat Theory (TST)* [59] and *Transition Path Sampling (TPS)* [60] simply partition the system into two and analyze the ensemble of transitions between them. These schemes, however are uninformative. The detailed trajectories may way complicated on the half way of transition. This can be illustrated by a maze example:

A walker starts from A and targeting B. In most cases, it will get lost on several dead ends on its way and may return to A before reaching B, those will be miscounted in TST. Even if we do focus on the path that do reach B before returning to A, the complex wandering during this process was missed. In TPT, it quantifies the probability flux cross any surface

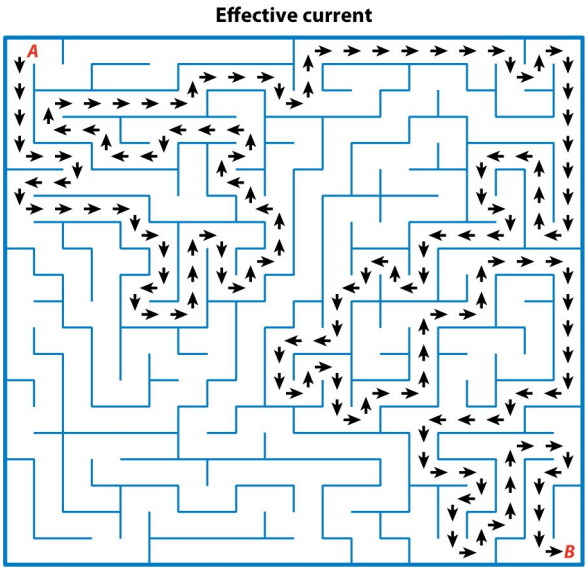


Figure 3.8.1: The maze example of the transition path [1]

on the half way of reaction by the so-called committor function $m(x)$ [61] [1]. We formally define it: *committor function is the probability that the trajectory starting from $x \notin A \cup B$ reaches first B rather than A.*

It's obvious that the probability density to observe a trajectory at x (whether reactive or not) is $\frac{e^{-\beta U(x)}}{Z}$ where Z is the partition function. The equivalence of distribution in trajectory ensemble and Boltzmann ensemble is guaranteed by ergodic hypothesis. Then conditional on that, one should times $m(x)(1-m(x))$ for a trajectory will be reactive. Summarizing, the probability density to observe a reactive trajectory at point $x \notin A \cup B$ at time spot t is:

$$\frac{e^{-\beta U(x)}}{Z} m(\mathbf{x})(1-m(\mathbf{x})) \tag{3.8.6}$$

then properties such as probability current, flux and reaction rate all can be further derived from that.

The implementation of TPT can be done in a "continuous" manner by binning the region between the reactant and product as a mesh [61], or it integrates well with MSM [62], the coarse-grained macrostates can be regarded as a discretized version of binning. As for raw data obtaining from simulations, for a simple equation of motion, one can obtain a

long trajectory (i.e. considerable transition events included), then prune this trajectory to the trajectory ensemble we want [61]; for complicated system, such as MD simulation of a macromolecule, the sampling process relies on numerical method such as *String Method* [63] to obtain reactive trajectories efficiently.

3.8.3 Connection between Importance Sampling and Enhanced Sampling

The view of trajectory space of sampling problem is actually connect the enhanced sampling and importance sampling by probability theory. The problem of importance sampling can be regarded as a optimization problem under constraints [57]. In which, one focus on a trajectory space Ξ, and h be a real valued function defined on Ξ with probability distribution p. The expectation value of h can be calculated by:

$$\mathbb{E}_p[h] = \int_{\Xi} h(x)p(dx) \tag{3.8.7}$$

However, mathematically, when p is an extreme nonuniform function, the samples are wasted around the maximum but may never visit the tailed regions. In MD simulation, it can be an analogues that among the whole trajectory space, the probability of observing a transition between metastable states will be small. In importance sampling, instead of draw samples from the extreme distribution p, one draw samples from a "proposal distribution q", such that the probability of transition events in this distribution are higher than those in original. Then, the samples can be drawn from the proposal distribution to make the sampling process more efficient. Finally, the quantities of interest can be re-weighted from the samples of q as:

$$\mathbb{E}_p[h] = \frac{1}{N}\sum_{i=1}^{N} h(\Xi^i)w(\Xi^i), \Xi^i \sim q \tag{3.8.8}$$

the w here is the weight of each value $h(\Xi^i)$ in the original distribution. The existence of the measure of re-weight is guaranteed by *Radon-Nikodym theorem*, that is the absolute continuity of p respect to q. One can check reference [64] and [57].

In molecular dynamics, the above approach is implemented by exploiting the Boltzmann distribution corresponding to the energy function. The "proposal distribution" in this case is the biased distribution after adding bias potential. In the context of the Variationally Enhanced method we mentioned in section 1.2, one can select a "proposal distribution" so does the bias potential [19]. In most cases, such as Metadynamics, Umbrella Sampling, the proposed distribution is a constant, in which the target is flat the energy surface.

We want this discussion to complete the idea of rare-event sampling and enrich the view of audiences. It should be recognized that the sampling techniques in multiscale simulation still have ample space to be developed. This could be achieved by keep pushing the accuracy and efficiency of choice of bias potential and re-weighting regimes from lending corresponded mathematical concepts or directly marry mathematical concept from math such as probability theory and stochastic process into simulation context. Below, we will use the MSM for our analysis of Brownian dynamics simulation, a preliminary result by TPT will also be presented at the end of this section.

3.9 Markov State Model and on/off rate estimation

Now we return to the numerical treatment of this problem. One can check the section 3.2 for the detailed setting. Recall that the model injects a particle from the equilibrium distance as a source, after the former one reacted on the sphere, allowing the system to reach a steady state. We will briefly extend the discusion of Nonequilibrium system later.

The aim of the current and following sections is using advanced statistical techniques to systematically interpret the simulated Langevin dynamics trajectory data. One could review section 3.8 for the detail of Markov State Model and Transition Path Theory. Here we give a brief reminder. On first divides a high dimensional set of data points into several clusters, each one denoting a distinct macro-state. One then use statistical techniques to investigate the kinetic network of those macrostates. Mathematically, they are applications of continuous time Markov Chain (CTMC).

Here, as our system already a coarse-grained description, it is straightforward that we can discretize the system into a four states model. Namely, four states are: the state unbound from (off) the sphere, bound to the sphere on the non-reactive region, bound to the sphere on the reactive region and the reaction state (as shown in figure 3.9.1). Here for the implementation, we use simple unsupervised learning, k-means algorithm for this discretization process.

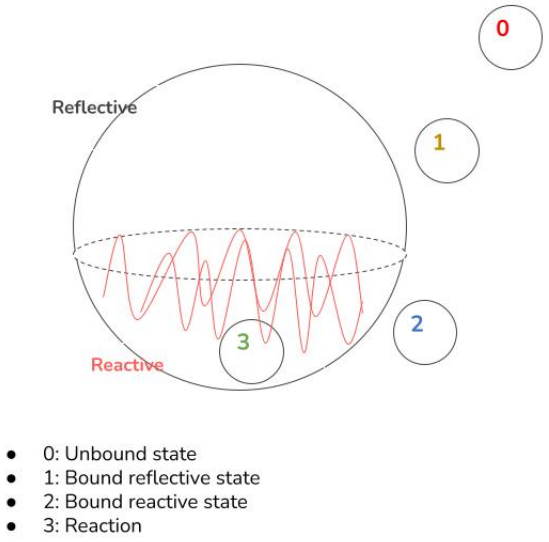


Figure 3.9.1: Cartoon representation of the 4-states model

Then, one can follow the section 3.8 to build a transition matrix. Noticed, to create a transition matrix or $\mathbf{Q}$ matrix of a continuous time Markov chain, one need to first have a count matrix $\mathbf{C}$, which entries are a function of lag time τ. The lag describes the time interval in which transitions between states are defined, a reasonable choice of τ should guarantee sufficient transitions between each state and makes the transition Markovian. Practically, one should check the implied timescales of the different processes and see that they don't change as a function of lag time to make the choice appropriate. The transition matrix is then constructed by dividing the the entries of each row by the summation of the row.

$$\mathbf{Q_{ij}} = \frac{\mathbf{C_{ij}}}{\sum_j \mathbf{C_{ij}}}$$

The transition matrix is the heart of MSM. After the matrix comes in handy, one can

learn a lot about the corresponding system by post-analyzing the matrix. First, one can use the solution of Kolmogrov equation:

$$\vec{P(t)} = e^{-t\mathbf{K}}\vec{P}_0 \tag{3.9.1}$$

to get the rate matrix. Then the current between paired states are:

$$J_{ij} = P_{ij}K_{ij} - P_{ji}K_{ji} \tag{3.9.2}$$

So does the on/off rate and net current through the system can be estimated. The dynamics of the system can also be learned by analysis of eigenvalue structure of the matrix. Finally, the transition path of mechanism can be investigated by calculating the "commitor function" and transition distribution. We will directly present the results of those applications in the following sections, and again encourage audiences to read the section 3.8 for the mathematical details.

3.10 Numerical estimation of surface concentration

By implementing Brownian Dynamics as mentioned in section 3.4, one could also get information of surface concentration for the steady-state system by simply counting number of particles appeared on the surface. The particles be counted are the one presented in the on state which has been defined by the discretization process mentioned above. Here, to be consistent with the analytical work, we present the case with the hemisphere reactive and reactivity $p = 0.5$. The resulted distribution has a good agreement with the analytical derived profile with finite evaporation rate.

We also traced the time evolution of the surface concentration distribution by counting the particles in different time interval. The resulted distribution change is similar with the theory, where the initial distribution is nearly flat and then particles are gradually accumulated in the nonreactive region as shown in 3.10.1

Notice that the density on the reactive region changes monotonically in this numerical result, which does not agree with the theory. We speculate that this due to the different

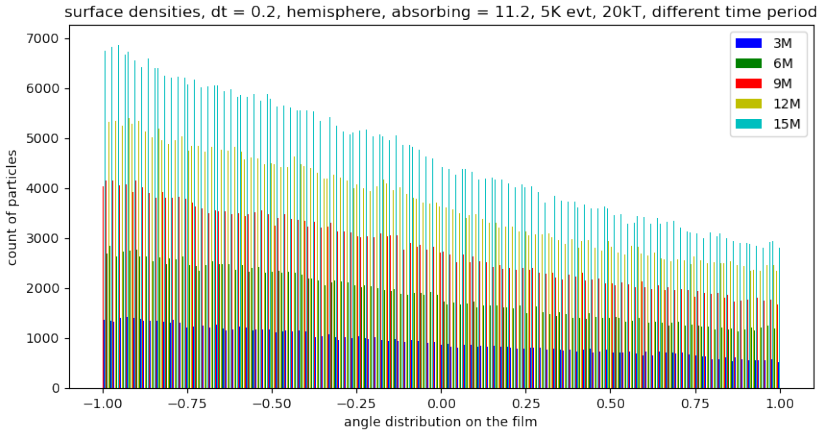


Figure 3.10.1: Time series of surface concentration, each histogram distinct a distribution at different time spots

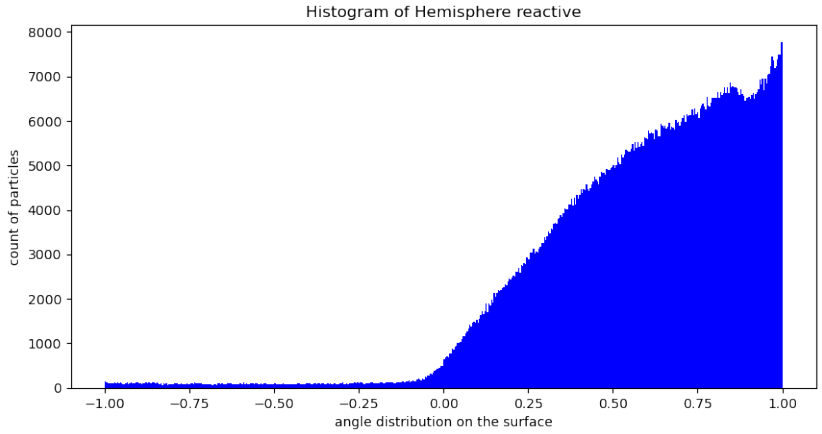


Figure 3.10.2: Steady state surface concentration of the hemisphere case

deposition mechanisms between the two models. We will discuss this in detail in the following section of reaction rate.

3.11 First hitting time and reaction rate

The relation between reaction rate and area of reactive region (open angles) is the main result of this project. We already calculated the reaction rate analytically as the incoming current which cross the reactive/reflective boundary. Here we use two approaches, mean first

passage time (MFPT) and net current through the system by MSM to numercically estimate the reaction rate. The discrepancies between numerical results and theory will be discussed. The inspiration to biological target searching process will also be reviewed.

We first investigated the reaction rate by 1 over the expectation of first hitting time or MFPT, $\frac{1}{\langle t \rangle} = \frac{1}{MFPT}$ or $\frac{N_t}{T}$ where N_t is the number of reaction events and T is the total length of the trajectory. This property can be also explained as the total current through the system as we mentioned earlier. The implementation of the first hitting time calculation is exactly identical to the simulation of steady state distribution in section 3.4. Here we generate more events (eject more particles), i.e. 5000 events for each calculation. The parameters such as spring constant $k = \frac{1}{600}$ and the reactivity $p = 0.5$ for the Langevin dynamics simulation were selected. For $k = \frac{1}{600}$ there was a good confinement of the particle but its corresponding equilibrium distance are out of L-J's effective range, thus one expected to observe the competition between 3-dimensional diffusion and 2-dimensional diffusion which enhanced by L-J. As for reactivity, here we use $p = 0.5$ as an intermediate and simple case.

To be consistent with analytical works, we first checked the rate without L-J attraction, that is the particle's displacement are governed only by harmonic potential and stochastic force. Notice that simulation of the particle is confined to a certain range. This is different from the setting of analytical solution, in which the particle diffused in from infinity. However, the resulting reaction rate is nearly linear with respect to the opening angles of the reactive part of the sphere. This means that the Berg-Purcell model still is valid. Quantitative agreement and discrepancy between the analytic and numeric models will be discussed later

Next, we turned on the L-J potential. To compare with the analytic solution with non-zero evaporation rate, we also implemented several simulations with different strength of L-J. There is a significant difference between L-J and evaporation in analytical model. In the latter case, evaporation corresponds to the rate of escape out of the LJ well. Larger evaporation rates correspond to weaker bound state so weaker L-J potentials in the numerical simulation. The dependence of the reaction rate on different open angles is presented in

3.11.1:

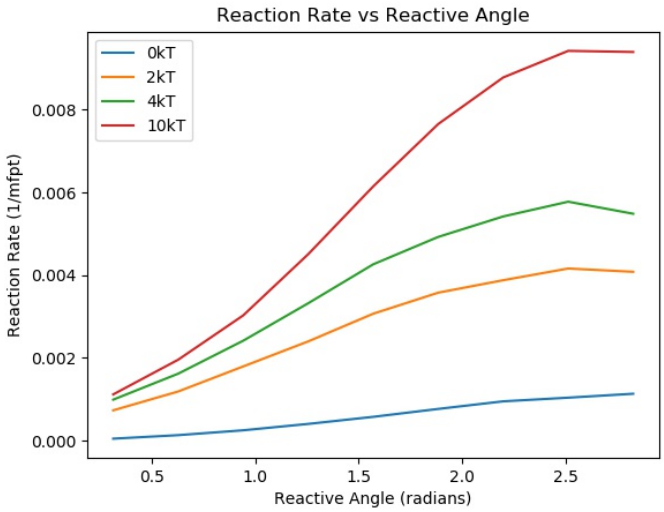


Figure 3.11.1: Reaction rates respect to different open angles (each curve) with different L-J potential, calculated by MFPT

The blue curve is the pure-diffusive case without L-J as mentioned earlier in this section. For the three curves with different L-J, they all monotonically increase respect to open angles and emerge a positive curvature and converged with large angles which we saw in analytical solution. The linearity at small angle limit implies the Berg-Purcell model still holds for the case with short range interaction, i.e. L-J potential.

Different from the analytical solution, however, we do not observe the crossover between curves with different strength of L-J. The numerical results presented here do indicate the rate proportional to the depth of L-J for all angles, but no optimal condition exits. This observation conflict with the one we obtained from section 3.6, in which increased attraction did not always mean increased reaction rates. Specifically, for the case of small angles, intermediate binding strength can produce a higher reaction rate. A deeper potential well results in a longer two dimensional searching process if the reactive region is small enough. For intermediate depth of the potential, the particle gets the chance to hop off the surface and then return to the surface to start a fresh search, which speeds up the reaction. From this perspective, we could also explain the discrepancy between analytic and numeric results.

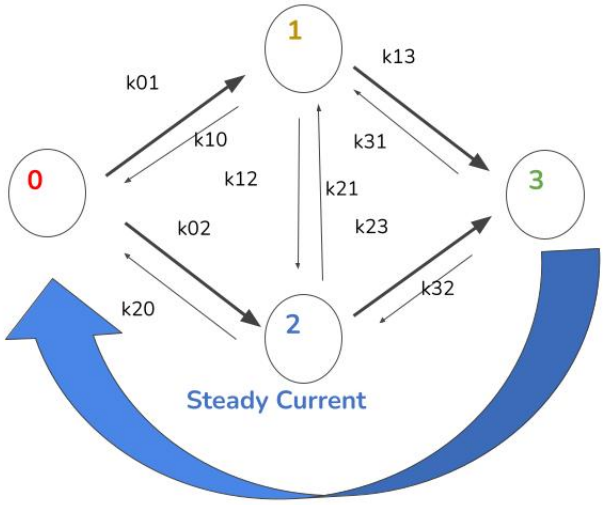


Figure 3.11.2: Cartoon representation of 4-states model and currents through it

In the analytical model, one assumes that the particles are uniformly redistributed over the whole sphere after evaporation. The three-dimensional searching method would be helpful in this case. Differently,in the numerical simulation we have a short range interaction, in which the particle will be locally redistributed to its vicinity but not really an efficient three-dimensional relocation. Along with this, a particle which does escape may also take a very long time to return to the surface.

We confirmed our result by calculating the current through by MSM analysis. In the four-states model we created, the current from the limbo state, that is the reaction state to the source is essentially the reaction rate. Theoretically, in the non-equilibrium steady state, such current is also the net current through the system, i.e. the net flux across each state. For a specific state i, this could be calculated by:

$$J_i = \sum_{j \neq i} P_i k_{ij} - P_j k_{ji} \tag{3.11.1}$$

where P_i and k_{ij} are stationary distribution and transition rate which came in handy when we obtained the transition matrix.

The corresponded kinetic scheme can be summarized by the cartoon below.

The master equation is then:

$$\frac{dP_0}{dt} = -(k_{01} + k_{02})P_0 + k_{10}P_1 + k_{20}P_2 + \frac{dP_3}{dt} \tag{3.11.2}$$

$$\frac{dP_1}{dt} = -(k_{10} + k_{12} + k_{13})P_1 + k_{01}P_0 + k_{21}P_2 \tag{3.11.3}$$

$$\frac{dP_2}{dt} = -(k_{20} + k_{21} + k_{23})P_2 + k_{12}P_1 + k_{02}P_2 \tag{3.11.4}$$

$$\frac{dP_3}{dt} = k_{13}P_1 + k_{23}P_2 \tag{3.11.5}$$

then the resulted rate matrix is:

$$\begin{bmatrix} -(k_{01} + k_{02}) & k_{10} + k_{13} & k_{20} + k_{23} & 0 \\ k_{01} & -(k_{10} + k_{12} + k_{13}) & k_{21} & 0 \\ k_{02} & k_{12} & -(k_{20} + k_{21} + k_{23}) & 0 \\ 0 & k_{13} & k_{23} & 0 \end{bmatrix}$$

the extra terms in the first row for the off diagonal elements come from replacing $\frac{dP_3}{dt}$ in equation 3.11.2 with equation 3.11.5, as we just repeated, the state 3 is a sink where we set an artificial current from state 3 to state 0. Beyond that, since for this case the fourth column is all zeros due to the sink in state 3, we can ignore the fourth row and column and analyze the 3x3 submatrix instead. This will facilitate us in the process of taking the matrix log to solve the Kolmogrov forward and backward equation. The logorithm of transition matrix can be processed easier by diagonalizing it. Below is the computed rate respect to different open angles with L-J equals $10k_BT$.

The magnitude of net current has a good agreement with most of open angles. However, when the open angle large enough, current from state 2 to 0 was computed inaccurate due to oversimplification of the system, this lead to a dramatic decreasing of rate at large angles. One need to reconsider the selection of number of macrostates, but here we first stick with it but postpone the discussion later. The resulted eigenvalue structures of this three states system also illustrate the dynamics of the system, we discuss this in the following section.

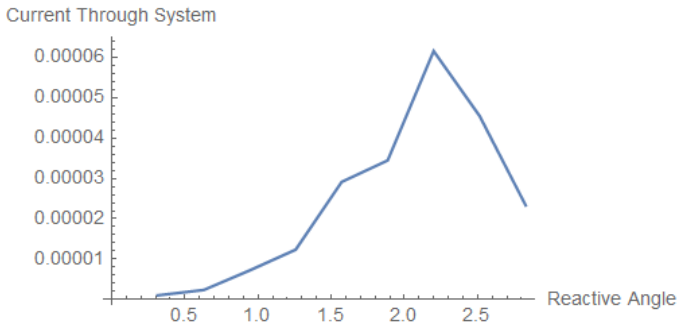


Figure 3.11.3: Net current through the system from 3 states model

The net current in equation 3.11.1 above is also proportional to entropy production rate of the non-equilibrium system, we will discuss more of this in the section 3.13 along with more non-equilibrium essentials of the system.

In a review by Mirny in 2008, the combination of 3-D diffusion and 2-D diffusion was discussed [65]. Most of the biological processes are controlled by diffusion limited reaction, it required reactant travel across the human cell, with a crowding molecule environment, the transport even slower as a subdiffusion process. However, the reactant molecules are actually faster than image to reach their target. In 1968, Delbruck [65] showed that the reduction of dimension could speed up the target searching process. A typical example is the protein-DNA binding problem, in that situation, the protein first binds DNA in a random location, then do a 1D sliding to search the targeted fragment. Different from simple 1D searching, the pure 2D process sometimes redundant, where the same area will be visited repeatedly by a small excursion, thus the searching is inefficient. Loverdo et.al [66]. argued the combination of 2D and 3D diffusion is a good strategy. The long distance jump helps a particle to explore new locations, thus making the search less redundant and faster. This optimal condition was also observed in our theory, in which the hopped particle has chance to relocate anywhere on the sphere as we set it can deposit on the surface uniformly. The discrepancy between theory and numerical model can be explained the hopping with a short-range interaction cannot lead to a fast transport in 3D but a excursion around its hopping location.

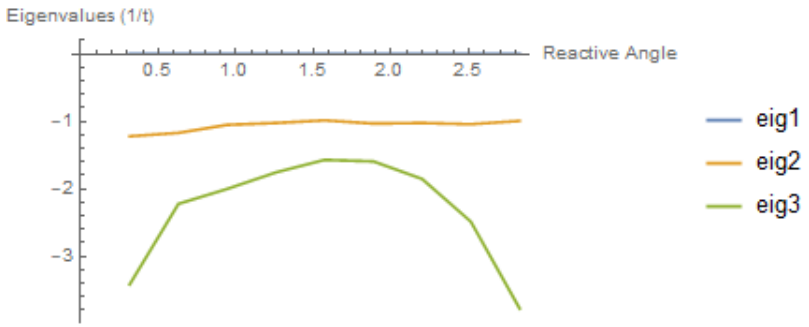


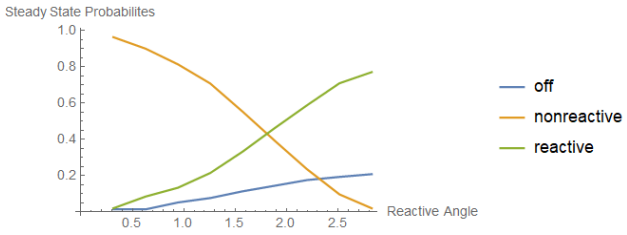
Figure 3.12.1: Three lowest eigenvalues as a function of the opening angle.

3.12 Reaction dynamics by MSM analysis

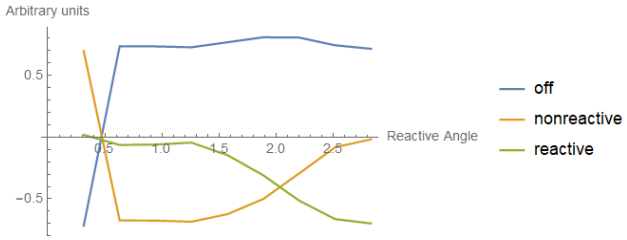
As we briefly mentioned at section 3.7, we could understand the dynamics of the system before reaching the final steady state. Similar with the homogeneous linear equation, we can analyze the eigenvalue structures of the transition matrix to investigate more about the dynamic process.

The first two eigenvalues with respect to different angles have a good agreement with the theoretical prediction. The first eigenvalue corresponds to the steady state that with all open angles it emerges zero decay rate. The second and the third one from analytical model are constant decay rate which denoting the thermal fluctuations before reaching the final state are gradually vanished. The second modes was also observed from the numerical analysis. However, the third mode shows non-monotonic behavior with respect to the opening angles. Specifically for the two extremes with large and small opening angles, the decay rates are very large. This discrepancy in eigenvalue structures resulted from the limited basis set in the numerical model. The coarse grained four states model is not comparable with the (roughly) one hundred terms of the Legendre polynomial expansion.

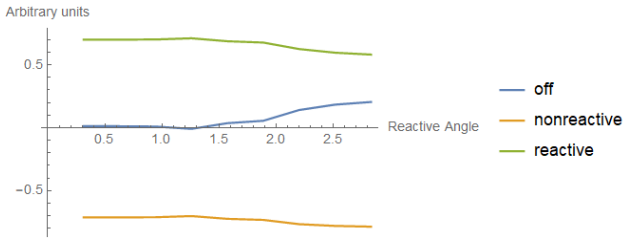
These trends were also appeared in the eigenvectors which corresponds to the probability distributions of each mode. Figure 3.12.2 gives the relative probability of each mode with respect to open angles. The steady state distribution changed monotonically with increasing angle. The second and the third eigenvalues indicate a shift of mechanism when we increase the opening angle.



(a) steady state



(b) 2^{nd} mode



(c) 3^{rd} mode

Figure 3.12.2: Eigenvectors of three modes

3.13 Entropy production rate

In this section, we will discuss more properties of the system in macroscopic view. As we keep generate particles from source, the process is clearly a steady state non-equilibrium system. By thermodynamics meaning, it is irreversible and will keep producing entropy. This essential violate the detailed balance which is the characteristic of an equilibrium system. This violation could be checked by Kolmogrov criterion: $k_{ij}k_{jk}k_{ki} \neq k_{ik}k_{kj}k_{ji}$ [67].

Different from master equation and transition matrix which illustrated the process microscopically, entropy is a macroscopic description. To bridge the gap between micro and macroscopic description, we define a quantity, entropy production [68]:

$$\dot{S} = \frac{1}{2}(\sum_{i \neq j} P_i k_{ij} - P_j k_{ji}) ln(\frac{k_{ij}}{k_{ji}}) \tag{3.13.1}$$

From a thermodynamic point of view, this quantity manifest the system's response to the coupled external "force" [69]. In our case, the force are given as a net flux due to reaction and continuous fed of reactant (injection of particle).

The accurate calculation of entropy production rate requires we include the limbo state, cause in the 3 state model the rate matrix was simplified to

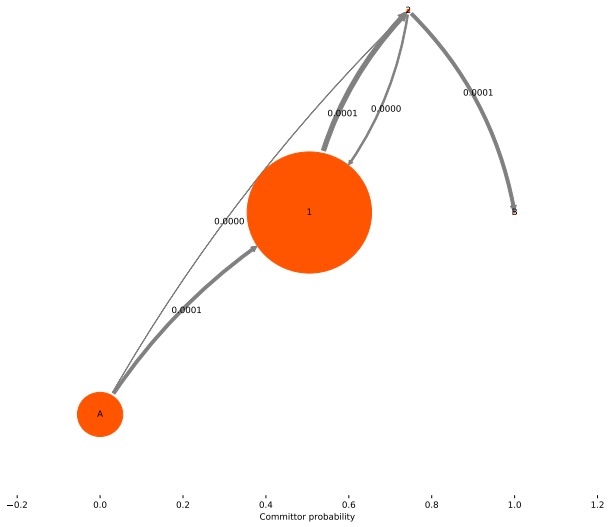
$$\begin{bmatrix} -(k_{01}+k_{02}) & k_{10}* & k_{20}* \\ k_{01} & -(k_{10}*+k_{12}) & k_{21} \\ k_{02} & k_{12} & -(k_{20}*+k_{21}) \end{bmatrix}$$

here k_{13} and k_{23} have been eliminated from the matrix and the k_{10} and k_{20} rates have been modified. The rates from states 1 and 2 to 3 have been eliminated since now we are no longer counting the number of reactions happening from state 1 and 2 to 3. They are instead counted as transition from the bound states, 1 and 2, to the unbound state 0. This reorganization is a good approximation when the hopping rate is ignorable, and the transition from state 1 & 2 to 3 so does 0 are dominant. However, ignoring the hopping transition would make the estimation of entropy production inaccurate.

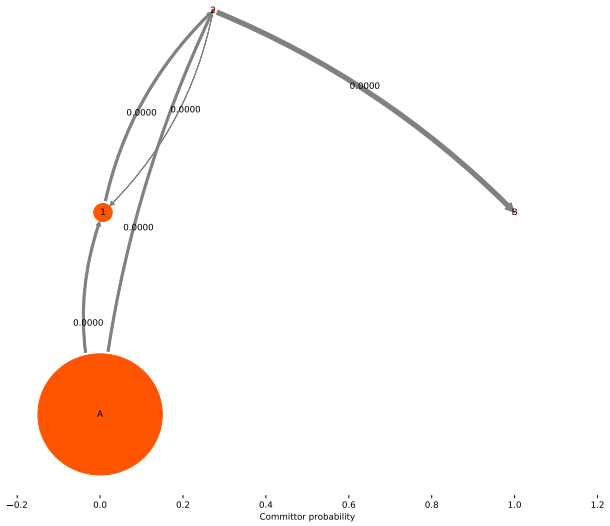
The truncation of four state matrix is not trivial, where every non-diagonal element is non-negative, which is known as the embedding problem. This problem come from *Galois theory* is way profound in our discussion, we here refer to [70] for a possible approximation.

3.14 Transition path and reaction mechanism

We take two extreme cases of open angles for the TPT analysis to illustrate the difference of open angles that will shift the transition mechanisms. As the model is a coarse-grained description, we use the transition matrix to calculate the probability distribution of the transition path by a sequence of conditional probabilities along the different possible paths from initial state 0 or off state to the final state 3 or the reaction. Resulted plots are shown in 3.14.1. The horizontal axis here is the probability that starts from 0 or A and ends in 3 or B, known as committor probability. The size of the circle indicates the net flux through the state. One can easily see, for a small angle case $\frac{\pi}{10}$, the flux was dominant at the reflective state where the particles accumulated there for the target searching process. On the contrary, with a large angle, the reactant state is dominant as almost all particles directly rain down to the reactive spot without any stay.



(a) committor probability calculated from the simulation with $\frac{\pi}{10}$ open angle



(b) committor probability calculated from the simulation with $\frac{9\pi}{10}$ open angle

Figure 3.14.1: Committor probabilities with two open angles

CHAPTER 4

Growing a Droplet in a Filament Network

4.1 Liquid-liquid phase separation in living cells

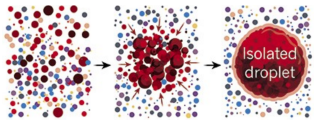
Liquid-liquid phase separation occurs when a change in thermodynamic parameters causes the components of a binary mixture to segregate. Liquid-liquid phase separation eventually produces an interface separating two components, such as oil floating on water. Below is an artistic interpretation of phase separation inside the complex environment of a cell.

This phenomenon is common in biological systems and it may play a purposeful role as a mechanism that allows cells to respond to the changes in the environment. By concentrating certain compounds and excluding others, it can speed up the reaction [71]. For example, in neurons signaling molecules may aggregate to ensure smooth communication [72]; Phase separation also can protect cells by helping them to adapt to the environmental change, such as dramatic temperature and pH change [73]. However, phase separation could also damage the cell if it occurs at the wrong place or time, by creating clogs or aggregates of molecules which correspond to neurodegenerative diseases, such as ALS [74]. The figure below shows examples.

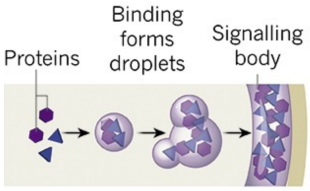
Different from typical liquid-liquid phase separation, the phase separation in the cellular system usually inter-plays with cross-linked polymer networks, which makes it more complicated.

Past work focused on the elastic resistance of the medium against the swelling pressure of the growing droplet. One work that should be mentioned here is the experimental work from the Dufresne group [75]. In their work, they argued that the growth of the droplet is driven by

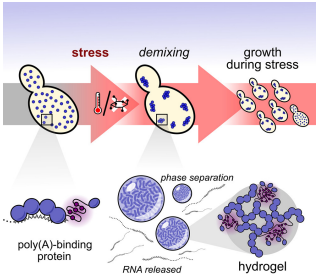
Figure 4.1.1: Artistic interpretation of phase separation inside a cell.



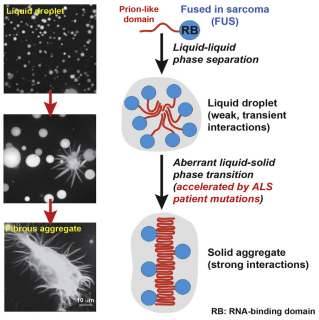
(a) speed up reactions [71]



(b) ensure a smooth communication [72]



(c) protect cells [73]



(d) neurodegenerative disease [74]

Figure 4.1.2: Examples of phase separation in cells

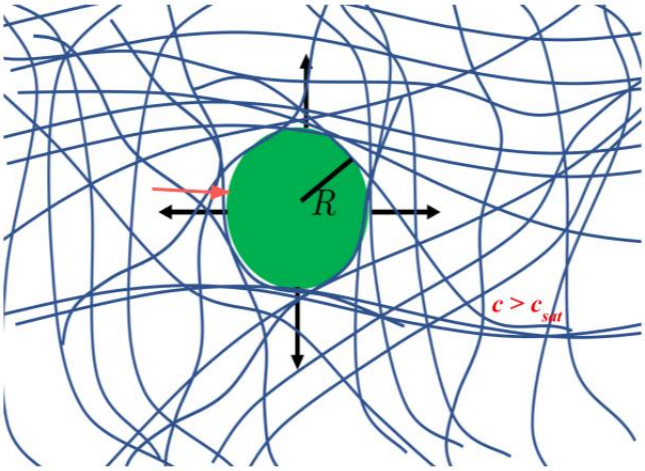


Figure 4.1.3: Phase separation according to linear elastic theory

the chemical potential difference between the dilute minority component molecules at infinity and the dense minority component inside the droplet. In their model, the swelling droplet keeps interacting mechanically with the elastic medium. As the droplet swells, cavitation phenomena may happen when linear elasticity breaks down.

Recent theoretical work focused on the impact of stiffness & shear modulus of surrounding networks on the phase separation phenomenon. The Zwicker group predicted that the distribution of the droplets is a result of the gradient of stiffness and can be regarded as a reversed Ostwald ripening [76][77]. The Meng group verified experimental work of the proportionality between droplet size, number density, and shear modulus of the elastic network [78]. All works, however, modeled the system by using mean field theory and the interaction between droplet and network were described by continuum mechanics.

4.2 Two dimensional model of the growing droplet

4.2.1 Basic Settings

Different from past works we reviewed above, we create a two-dimensional model dynamical model for the numerical simulation. In this model, a hexagon lattice of deformable Gaussian potentials is used to represent either filaments or a coarse-grained representation of a higher filament concentration region. This allows one to simulate different elastic networks by changing parameters of the Gaussian potentials. Harmonic spring connecting the Gaussians provides mechanical stability of the network. Next, a disk is used to represent a droplet of minority phase. The motion of the Gaussian bumps and the droplet are both governed by Brownian dynamics with external potentials:

$$\dot{X}_d(t) = -\frac{D_d}{k_B T}\nabla U_{int} + \sqrt{2D_d}\sigma(t) \tag{4.2.1}$$

$$\dot{X}_g(t) = -\frac{D_g}{k_B T}(-\nabla U_{int} + \nabla U_h) + \sqrt{2D_g}\sigma(t) \tag{4.2.2}$$

Here, X_d and X_g are locations, D_d and D_g are the diffusion constants of the droplet and Gaussians respectively while σ again is the Wiener noise we used previously. U_h is the harmonic potentials between each pair of Gaussians:

$$U_h = \frac{1}{2}k(\vec{x_{gi}} - \vec{x_{gj}})^2 \tag{4.2.3}$$

where $\vec{x_{gi}}$ and $\vec{x_{gj}}$ are locations of the Gaussian i and j.

Importantly, the droplet has an interaction potential U_{int} that depends on the overlapping area between the droplet and Gaussians. Conversely, a Gaussian will interact with the droplet by Newton's third law, for which explains the negative sign in equation 31. Importantly, the size of the droplet is a variable that depends on U_{int}, surface tension, chemical potential, and the dewetting parameter. The dynamics of radius (size of the droplet) will couple with the

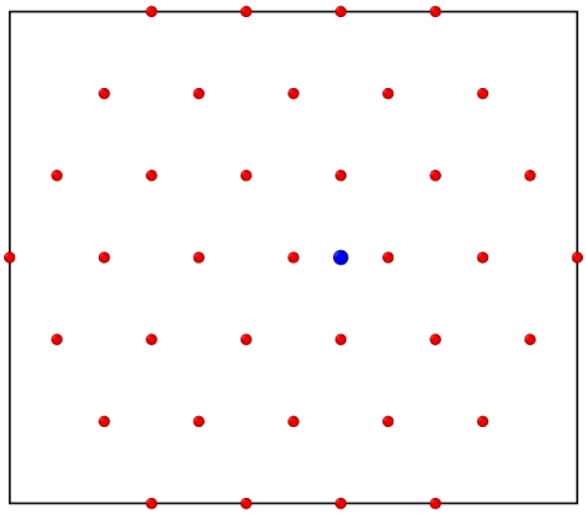


Figure 4.2.1: 2D lattice of Gaussian (red dots) with a droplet (blue dot)

motions of the droplet and Gaussians, which leads to intriguing properties. I will explain this in the next section.

4.3 Interaction between Gaussians and growing droplet

4.3.1 Overlapping & Interaction Energy

As we just mentioned, the interaction energy has a key impact on the dynamics of the system. We now formally define how it depends on the overlap between droplet and Gaussian and then how to calculate it use a "soft droplet" trick.

First we define the magnitude of Gaussian potential density. For an arbitrary Gaussian with its center locate at $\vec{x_g}$, the potential U_g is:

$$U_g(\vec{X}) = U_0 e^{-\frac{1}{2\sigma^2}(\vec{X}-\vec{x_g})^2} \tag{4.3.1}$$

Here, U_0 is the maximum of Gaussian and σ is the standard deviation. Those two parameters represent the height and width of the bump, it allows one to mimic different strength & density of filaments. We further set a disk with radius R at $\vec{x}$, and $\vec{y}$ is the vector locate

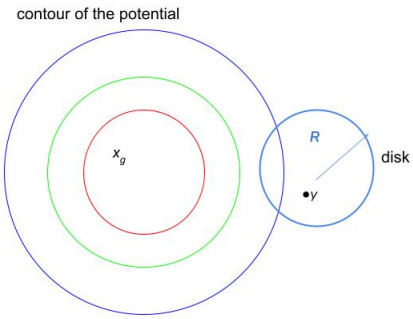


Figure 4.3.1: Cartoon representation of the interaction between the Gaussian and the droplet

within the disk, it is the substitution of an arbitrary vector respect to the center of Gaussian $\vec{X} - \vec{x_g}$. The interaction energy is the integral of the potential density over the area of the disk. A naive cartoon is shown below:

Intuitively, we use this interaction potential to simulation the forces generated by filaments' deformation. Mathematically it follows:

$$U_{int} = \int_{|\vec{y}-\vec{x}<R|} U_g(\vec{y}) d^2\vec{y} \tag{4.3.2}$$

Normally we set the disk as a "hard disk" with sharp edge, i.e. the vector $\vec{y}$ are limited within the edge. Then we can rewrite the equation 4.3.2 into a polar coordinate form:

$$U_{int} = \int_0^{2\pi} d\theta \int_0^R r dr e^{-\frac{x^2+2xrcos\theta+r^2}{2\sigma^2}} \tag{4.3.3}$$

Here we extend equation 4.3.2 by equation 4.3.1. The $\vec{x}$ would still be the location vector of the disk, however the vector $\vec{X} - \vec{x_g}$ in equation 4.3.1 be expressed in the polar coordinates as a vector $\vec{z}$:

$$\vec{z} = \vec{x} + \hat{x}rcos\theta + \hat{y}rsin\theta, 0 < r < R \tag{4.3.4}$$

A simple illustration is helpful for understanding the model:

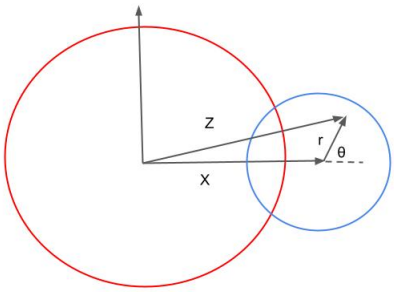


Figure 4.3.2: Toy representation of integral in polar coordinate

However, it is hard to truncate this integral analytically which requires Bessel function. We will put the derivation into the appendix session. Here, we regard the droplet as a "soft disk" to simplify the integral. In the "soft disk" version, we regard the disk itself as a Gaussian density:

$$\rho_D^{soft} = \rho_0 e^{-\frac{r^2}{2R^2}} \tag{4.3.5}$$

To be clear, we can also formulate the 'hard disk' we derived above, its density should be a step function:

$$\rho_D^{hard} = H(R - |r|) \tag{4.3.6}$$

and in equation 4.3.2, we drop the constant density $\rho_D^{hard} = 1$ within the integral range. Differently from the soft version, the part be integrated is fixed to be one which fulfills the condition that the minority phase (droplet) has a fixed amount. With this setting, the interaction energy will be:

$$U_{int} = \int d^2\vec{y} U_g(\vec{y})\rho(\vec{y}) \tag{4.3.7}$$

Along with equation 4.3.1, the integral would be:

$$U_{int} = U_0 \int d^2\vec{y} e^{-\frac{\vec{y}^2}{2\sigma^2}} e^{-\frac{|\vec{y}-\vec{x}|^2}{2R^2}} \tag{4.3.8}$$

A simple illustration would be exactly identical as before. Here we first focus on the argument of exponential term, it can be collected and then reorganized by completing the square:

$$-\frac{1}{2\sigma^2}y^2 - \frac{1}{2R^2}(y^2 - 2\vec{x}\vec{y} + \vec{x}^2) = -\frac{x^2}{2(\sigma^2+R^2)} + \frac{R^2+\sigma^2}{2\sigma^2R^2}(\vec{y} - \frac{\sigma^2}{R^2+\sigma^2}\vec{x})^2 \tag{4.3.9}$$

plug it back with putting $\vec{y}$ irrelevant term in front of the integral:

$$U_{int} = e^{-\frac{x^2}{2(\sigma^2+R^2)}} \int d^2\vec{y} e^{\frac{R^2+\sigma^2}{2\sigma^2R^2}(\vec{y}-\frac{\sigma^2}{R^2+\sigma^2}\vec{x})^2} \tag{4.3.10}$$

which is a Gaussian integral and the result is straightforward:

$$U_{int} = \pi(\frac{2\sigma^2}{R^2+\sigma^2})U_0 e^{-\frac{x^2}{2(\sigma^2+R^2)}} \tag{4.3.11}$$

The resulting expression shows that the interaction strength grows with increasing R, which is consistent with intuition.

4.3.2 Chemical Interaction

We now start to formulate the dynamics of R. We first focus on how a Gaussian interacts with er droplet chemically. For this purpose, we start from fundamental thermodynamics, where we assume the minority phase act like an ideal solution and its Gibbs free energy is:

$$F = Nk_BTln(\frac{n}{n_0 e}) \tag{4.3.12}$$

in which n is the number density of the minority phase, and $n_0 = (\frac{mk_BT}{2\pi\hbar^2})^{\frac{3}{2}}$ is a reference density. Then we can obtain the chemical potential:

$$\mu = \frac{\partial F}{\partial N}|_{T,V} = k_B T ln(\frac{n}{n_0}) \tag{4.3.13}$$

whew we applied the chain rule by using relation $n = \frac{N}{V}$. The gradient of the chemical potential will generate a current of particles $\vec{J} = -\nu \nabla \mu n$, where ν is the mobility. This relation is actually the diffusion equation:

$$\partial_t n = -\nabla \vec{J} = D \nabla^2 n \tag{4.3.14}$$

in two-dimensions. The solution will be:

$$n(r) = Alnr + B \tag{4.3.15}$$

where A and B are undetermined constants.

Now we focus on the equilibrium condition of the diffusion. In this condition the chemical potential of minority phase is balanced with that of mixed phase (ideal solution as we assumed): $\mu_{in} = \mu = k_B T ln(\frac{n^*}{n})$, in which n^* is an arbitrary number density of minority phase. Here we further set the number density at infinite region is a fixed value n_∞, which plays a role of a reservoir of minority phase. If those two regions relaxes separately as steady currents, we can get an equation series by plug them in equation 4.3.15:

$$n(R) = AlnR + B = n^*; n(\rho) = Aln\rho + B = n_\infty \tag{4.3.16}$$

here R is the radius of the droplet and ρ is the location at infinity. Then by solving this, the undetermined constants can be expressed by the constants we have:

$$A = \frac{n_\infty - n^*}{ln\frac{\rho}{R}}; B = \frac{n^*}{ln\frac{\rho}{R}} ln\rho - \frac{n_\infty lnR}{ln\frac{\rho}{R}} \tag{4.3.17}$$

now we plug them back to the equation 4.3.16, then take gradient respect to r. The constant B directly vanished and the current on the droplet R will be a simple expression:

$$J(R) = -\frac{D}{R}\frac{n_\infty - n^*}{ln\frac{\rho}{R}} \tag{4.3.18}$$

Now we can determine the dynamics of radius R by assuming the incoming current, the flux equation 4.3.18 times $2\pi R$, equals to the area change rate:

$$\frac{da}{dt} = 2\pi R\frac{D}{R}\frac{n_\infty - n^*}{ln\frac{\rho}{R}} = 2\pi R\dot{R} \tag{4.3.19}$$

where we applied $a = \pi R^2$. Then the radius change rate can be obtained by reorganizing 4.3.19:

$$\dot{R} = \frac{D}{R}\frac{n_\infty - n^*}{ln\frac{\rho}{R}} \tag{4.3.20}$$

then the number density of minority phase can be expressed by the equilibrium condition we mentioned above: $n^* = n_0 e^{\frac{\mu_{in}}{k_B T}}$. Since $\rho >> R$ is always true in our assumption, so we could further set $ln\frac{\rho}{R}$ as a constant $\tilde{D}$. Thus the radius change rate can be reorganized as:

$$\dot{R} = \frac{\tilde{D}}{R}(n_\infty - n_0 e^{\frac{\mu_{in}}{k_B T}}) \tag{4.3.21}$$

here the heart of our model is the chemical potential μ_{in} is dependent on the interaction energy:

$$\mu_{in}^{'} = \mu_{in} + \alpha\frac{U(\vec{x}, R)}{R^2} \tag{4.3.22}$$

one can regard the last term as the excess Gibbs free energy offered by interaction potential up to a parameter α. $\frac{1}{R^2}$ here represent unit area, alpha combining $\frac{1}{R^2}$ represent the number of molecules. So α is the density of minority phase, which is the intrinsic property of different material. In this equation it represents how sensitive the droplet against interactions, thus we name α as the "dewetting parameter". Noticed that, to make equation 4.3.22 reasonable, we assume that in the condensed phase the Gibbs free energy is nearly identical to Helmholtz free energy. We also drop the entropy term.

This equation give us good intuition about the dynamics of radius. It is controlled by a constant flow from infinity proportional to n_∞. When a droplet is located at the local minimum of the potential surface, the present chemical potential of the droplet $\mu^{'}$ is smaller than that at infinity, thus particles flows in so the droplet will grow. Conversely, the droplet free energy increases when it overlaps with the Gaussian, leading to a larger chemical potential than that in reservoir, thus particles flow out and the droplet shrinks. Below in figure 4.3.3 we again use a cartoon to illustrate this dynamics.

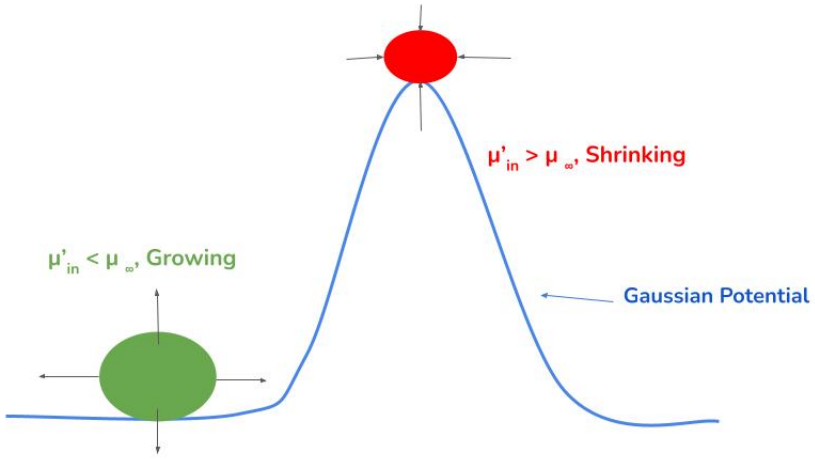


Figure 4.3.3: Cartoon of the size change of the droplet

Finally we could absorb all constant terms to a single parameter for a clean look, in which we assume:

$$\dot{R}_0 = \tilde{D} n_\infty; \dot{R}_1 = n_0 \tilde{D} e^{\frac{\mu_{in}}{k_B T}} \tag{4.3.23}$$

then equation 4.3.21 can be written as:

$$\dot{R} = \frac{1}{R}(\dot{R}_0 - \dot{R}_1 e^{\frac{\alpha U(\vec{x},R)}{k_B T R^2}}) \tag{4.3.24}$$

4.3.3 Mechanical Interaction

Besides the chemical interaction via excess Gibbs free energy, now we also considerate the mechanical interaction to the droplet. For which, we consider the surface/line tension of the droplet. We already know the chemical potential of the minority phase by equation 4.3.22, now if we add a small particle to the droplet, the updated chemical potential should be:

$$\mu^{'} = \mu + \frac{\Gamma a}{R} \tag{4.3.25}$$

in which $a = \frac{1}{n_0}$ is the area per molecule of the incoming cluster.

To prove equation 4.3.25, one just need to image a simple case: add a small particle to the drop, which result in a perturbation of radius, $R + dR$. Then we know the perimeter change should equals to: $2\pi dR$. Thus the area work (analogy of volume work in 2 dimension) should be:

$$\Delta\mu = \Gamma(2\pi dR) = \Gamma 2\pi \frac{a}{2\pi R} \tag{4.3.26}$$

where we use the relation that area change $2\pi R dR$ equals to the parameter a. As we assume the area a is the area/molecule, thus area work is the Gibbs free energy change per molecule, that is the chemical potential change. Then the chemical potential change from thermodynamic and mechanical interactions can be collected together, the resulted expression of $\dot{R}$ is:

$$\dot{R} = \frac{1}{R}(\dot{R}_0 - \dot{R}_1 e^{\frac{\alpha U(\vec{x},R)}{k_B T R^2} + \frac{\Gamma a}{R k_B T}}) \tag{4.3.27}$$

4.4 Summary.

We found that the parameter α controls the dynamical properties of the droplet. For α between six and ten, the droplet interacts strongly with the filament. The may evaporate shortly after the simulation start of the simulation or undergo a number of shrink-and-

growth cycles, producing the characteristic "shark fin shape" dependence of radius on time. Eventually, the droplet will evaporate when there is a sufficiently large thermal fluctuation. For α between 1.3 and 6, the system is in a quasi-equilibrium steady-state regime with the radius fluctuating around a mean value. Finally, for α smaller than 1.3, the droplet interacts only weakly with the surroundings. The radius will grow in time proportional to $t^{\frac{1}{2}}$, permeating the surrounding filaments. The changes that take place in the growth characteristics also affect the mobility of droplet. In both the evaporation regime and the steady-state regime, the droplet will be localized most of the time in a minimum energy location, with rare jumps to neighboring minima. In the diffusive regime of small α where the droplet permeates the surrounding filaments, the droplet will be much more mobile.

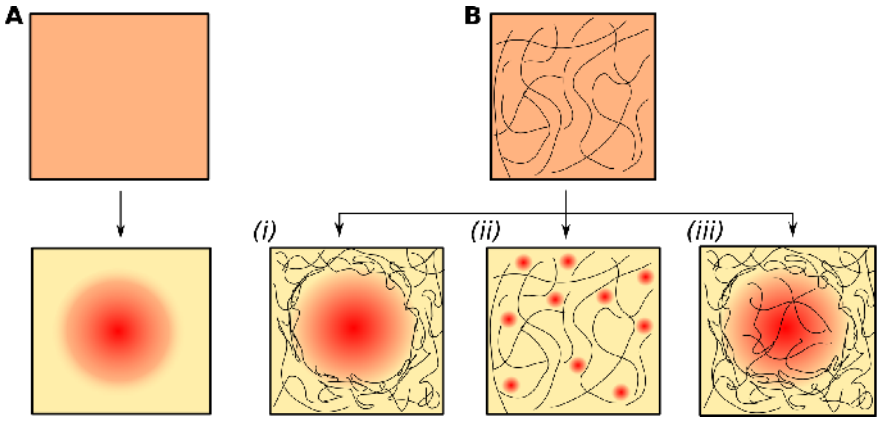
It is interesting to compare these regimes with the equilibrium phase diagram that was recently obtained by the Kosmrlj group for this problem [79]. The phase diagram obtained in that paper is shown in figure 4.4.1.

The regime of micro-droplets would correspond to the regime of large α parameter while the regime permeation corresponds to the regime of small α. The so-called permeo-elastic number is roughly inversely proportional to α. There is no analog in our theory for cavitation where the droplet pushes the filaments out of the way and it would be interesting to pursue that.

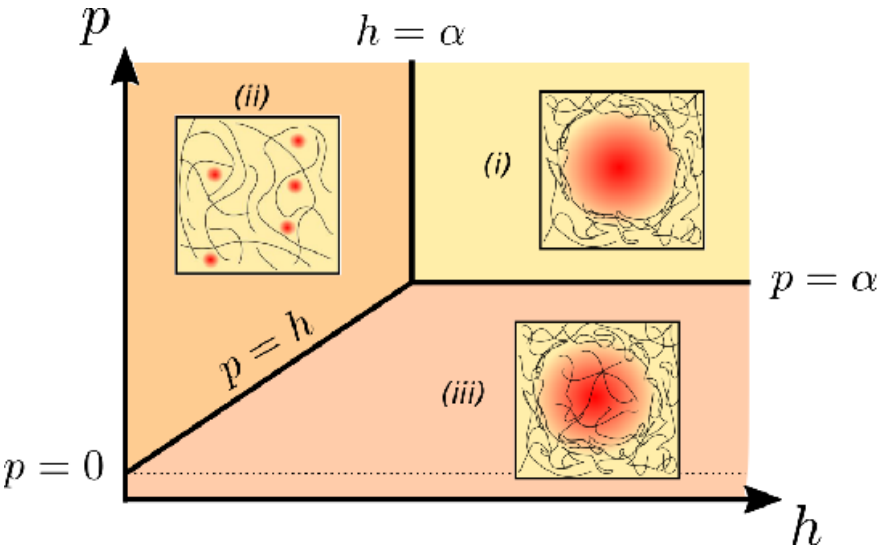
4.5 Radius fluctuations and concentration field

With our current settings, the fluxes are driven by the gradient of chemical potential, and the growth rate is identical to the incoming flux. This setting is under the assumption that the diffusion is large enough such that the concentration field along this transportation can be soon vanished, $\nabla^2 c = 0; \frac{\partial c}{\partial t} = 0$.

However, in actuality, the low and high density/concentration domains are separated by an interface. The relaxation of the concentration field is not instanteneous as we assumed. Instead of analytically solving the *Cahn-Hilliard equation* [80] for this diffusive system, we



(a) A. LLPS without interplaying with polymer networks; B. LLPS in an elastic network has three scenarios: (i) Cavitation; (ii) Micro-droplets; (iii) Permeation



(b) Phase diagram of LLPS in an elastic network as a function of elasto-capillary h and permeo-elastic number p

according to ref.[79]

Figure 4.4.1: Phase diagram and possible scenarios of LLPS

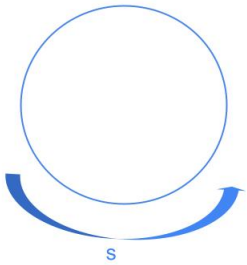


Figure 4.5.1: Toy representation of flux integral around the surface

use *Einstein relation* [81] to add a radius fluctuation to the dynamics of R as a proxy of the concentration field. Specifically, we add a noise term to the flux term which corresponds to the time evolution of the concentration profile:

$$\frac{\partial c}{\partial t} = -\vec{\nabla}\vec{J}; \vec{J} = -\nu c\vec{\nabla}\mu + \vec{J_n} \tag{4.5.1}$$

Again, here ν is the mobility and $\vec{J_n}$ is the noise term we add to the flux. Specifically, a path integral around the perimeter of the sphere was added to integrates fluxes around it:

$$\dot{R} = \frac{1}{R}(\dot{R_0} - \dot{R_1}e^{\frac{\alpha U(\vec{x},R)}{k_B T R^2} + \frac{\Gamma a}{R k_B T}}) + \frac{a}{2\pi}\int_0^{2\pi R} ds J_r^n \tag{4.5.2}$$

where s is the path along the perimeter and a again is the molecular area, and this integral can be represented as a cartoon in figure 4.5.1:

We now consider this correction term together as $\zeta(t) = \frac{a}{2\pi}\int_0^{2\pi R} ds J_r^n$. Then the auto-correlation function of $\zeta(t)$ is:

$$< \zeta(t)\zeta(t') >= (\frac{a}{2\pi})^2 \int_0^{2\pi R} ds \int_0^{2\pi R} ds' < J_r^n(s,t) J_r^n(s',t') > \tag{4.5.3}$$

it can be easily estimated as $2\pi R(\frac{a}{2\pi})^2 \Delta \frac{1}{l}\delta(t-t')$ up to an undetermined constant Δ. That relies on equal time correlation of Helmholtz free energy and kinetics of the concentration change: $\frac{\partial \delta c}{\partial t} = D\nabla^2 \delta c - \vec{\nabla}\vec{J_n}$. Here we only want audiences notice the radius would fluctuate

up to a noise due to such relaxation of the matter transfer. As we are not turning on this noise term in our later investigation, here We only leave the result is $\Delta = 2Dc(R)$.

4.6 Non-linear 1D system and Bifurcation

Before we turn on the fluctuations of dynamics of radius and displacement, we first simplified the system into a 1 dimensional model by coarse-graining the net force around droplet.

This simplification can be done by assuming several Gaussian potentials around the droplet exert an effective harmonic force on it, thus we could regard all point mass on the droplet feels such potential. For a location center of mass of the droplet $\vec{r}$ and an arbitrary point $\vec{\rho}$ deviates from the center, this integrated potential is:

$$U_{eff} = \int_{\vec{\rho}<|R|} d^2\vec{\rho}\frac{1}{2}u_0(\vec{r}+\vec{\rho})^2 = \frac{1}{2}u_0(r^2\pi R^2 + \frac{\pi}{2}R^4) \tag{4.6.1}$$

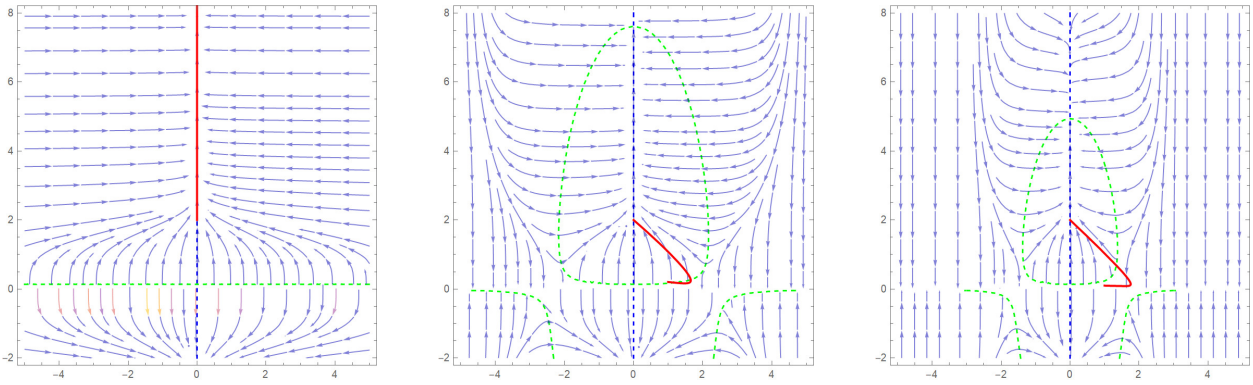
here u_0 is the effective spring constant of the harmonic, we measure this constant by simulating the 2D model with a small deviation, to make it comparable with the original model.

After obtaining the effective potential, we just need to replacing the U_{int} in equation 4.3.24 by this. The resulted expression has a clean look:

$$\dot{R} = \frac{1}{R}(\dot{R}_0 - \dot{R}_1 e^{\frac{\alpha u_0(r^2\pi+\frac{\pi}{2}R^2)}{k_B T}+\frac{\Gamma a}{Rk_B T}}) \tag{4.6.2}$$

Together with the equation of motion $\gamma\frac{d\vec{r}}{dt} = -\frac{\partial U}{\partial r}$ of the droplet, and with turning off both noise terms, the resulted differential equation series can be regarded as a nonlinear dynamical system. This deterministic description offers us insight in the dynamics. Here we use the resulted flow lines of vector field to illustrate it.

When we turn off the interaction energy, all vectors point towards the center location dominated by a regular harmonic potential. With $\alpha = 5$, a stable fixed point emerges at $r = 0, R = R^*$ in which R^* is the saturated radius, the horizontal vectors indicate at almost all finite displacement the radius tends to converge to the saturated radius. Two unstable



(a) $\alpha = 0$ (b) $\alpha = 5$ (c) $\alpha = 9.2$

Figure 4.6.1: Three regimes of flow lines

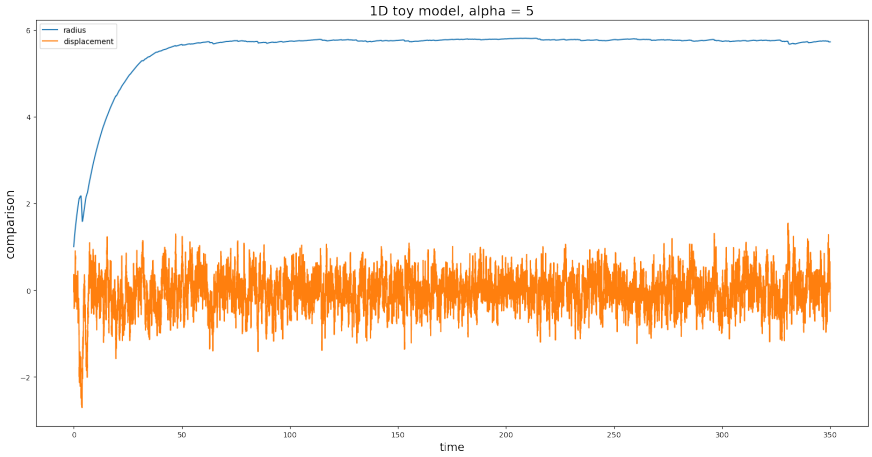
fixed points result from surface tension, when the size of the droplet is small enough, i.e. $R \approx 0$ the mechanical interaction forces the evaporation at a finite displacement. However, at later stage $\alpha = 9.2$, almost all vectors strictly rain down towards negative R, indicating the droplet will constantly evaporate even with a finite interaction with the potential.

4.7 The effective dynamics and theory by a perturbation

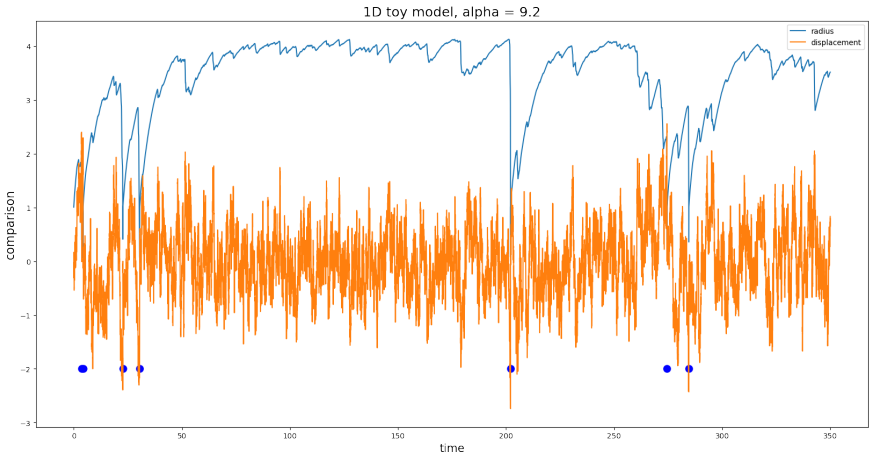
4.7.1 Stochastic dynamics

Now we turn on the thermal noise to use the stochastic dynamics to illustrate the difference between those two regimes. The resulted dynamics of radius and displacement with two different α are shown at top of this page:

With $\alpha = 5$, the radius rarely fluctuate but converge to the saturated radius. But with $\alpha = 9.2$, the evaporation events frequently happened which shown as blue dots in the below figures. In this case, the dissipation by interacting with potential cannot balanced by the thermal fluctuation, the restarting-fluctuation-evaporation process keeps consuming free energy and a steady state nonequilibrium condition has developed. In this case, the droplet cannot forget its initial condition with finite life time before it reaches the saturated radius state.



(a) $\alpha = 5$



(b) $\alpha = 9.2$

Figure 4.7.1: Two regimes of dynamics

4.7.2 A theoretical explanation by a small perturbation

The dynamics presented above is hard to truncate analytically. However, here we offer a theory to explain it by assume the radius change around saturated value up to a small perturbation: $R^2 = \bar{R}^2 + x(t)$, here $x(t)$ is the time-dependent perturbation. We further collect all prefactors as two constants in equation 4.6.1 in a cleaner form: $U_{eff} = \frac{1}{2}Cr^2R^2 + DR^4$. Thus at steady state, we have both differentiation equals zero, noticed here we also ignores the surface tension γ as the nonlinear dynamics showed, it just has light influence when the droplet small enough:

$$-CR^2r = 0; \bar{R} = \sqrt{\frac{1}{\alpha D} ln(\frac{\dot{R}_0}{\dot{R}_1})} \tag{4.7.1}$$

then the effective potential and the dynamics of the displacement can be written as:

$$U'_{eff} = \frac{1}{2}C(\bar{R}^2 + x)r^2 + D(\bar{R}^2 + x)^2; \zeta\frac{dr}{dt} = -C(\bar{R}^2 + x) + \eta(t) \tag{4.7.2}$$

here $\eta(t)$ is the white noise of the 1D dynamics. For $x = 0$, this is a trivial over-damped harmonic oscillator and the time correlation of the displacement has the form $< r(t)r(0) >=< r^2 > e^{-\frac{t}{\tau_r}}$ which up to a constant $\tau_r = \frac{\eta}{C\bar{R}^2}$. then the dynamics of x can be captured by plugging in the perturbation expression of R:

$$\frac{1}{2}\frac{dx}{dt} = \dot{R}_0 - \dot{R}_1 e^{\alpha(D\bar{R}^2 + \frac{1}{2}Cr^2 + Dx)} \tag{4.7.3}$$

by extracting constant term with $\bar{R}$ and D, one can Taylor expand the remaining exponential term as both r and x are small values at steady state. The resulting expression is:

$$\frac{1}{2}\frac{dx}{dt} = \dot{R}_0 - \dot{R}_1 e^{\alpha D\bar{R}^2}(1 + \frac{1}{2}\alpha Cr^2 + \alpha Dx) \tag{4.7.4}$$

By applying the saturated radius, equation 4.7.1, the above expression was cleaned up:

$$\frac{1}{2}\frac{dx}{dt} = -\dot{R}_0(\frac{1}{2}\alpha Cr^2 + \alpha Dx) \tag{4.7.5}$$

Next, we defined the time scale $\tau_R = \frac{1}{2\dot{R}_0 \alpha D}$, then we again rewrite the dynamics as:

$$\frac{dx}{dt} = -\frac{1}{\tau_R}x - \dot{R}_0 \alpha C r^2(t) \tag{4.7.6}$$

The dynamics of x can be solved analytically up to a integral:

$$x(t) = -\bar{R}_0 \alpha C \int_0^t dt' e^{-\frac{t-t'}{\tau_R}} r^2(t') + x(0) e^{-\frac{t}{\tau_R}} \tag{4.7.7}$$

then with a long timescale $t >> \tau_R$, the last term can be ignored. Then we plug in equation 4.7.7 into equation 4.7.2:

$$\zeta \frac{dr}{dt} = -C\bar{R}^2 r(t) + \bar{R}_0 \alpha C \int_0^t dt' e^{-\frac{t-t'}{\tau_R}} r^2(t') + \eta(t) \tag{4.7.8}$$

it is clear that the second term introduces a "memory kernel" to the dynamics of r, which consistent to the stochastic dynamics result, in which the droplet cannot forgot its initial condition with a finite life time before reach the equilibrium steady state. This non-uniformity in time is a strong evidence for a non-equilibrium condition.

4.8 Violation of fluctuation-dissipation theorem (FDT)

4.8.1 Non-equilibrium condition

Important biological processes such as the phase separation problem we now discussing consume energy, thus push the system away from equilibrium. Other examples such as molecular motors [82], proteins transport cargo [83] are now hot topics under intense research. In a wide range of such situations, one can categorize the non-equilibrium system into these three classes [84]. One is the system is slow relaxing towards the equilibrium but yet reached; the second is the system imposes a non-zero steady current through boundary condition (i.e., the constant particles we generated in the Diffusion confinement problem and the molecule current from a reservoir we discussing now); the third class of non-equilibrium systems is so-called active matter in which energy is dissipated at the microscopic scale in the bulk

so that each component of the system has an irreversible dynamic. In general, the non-equilibrium conditions are generated due to the non-uniformity in space or time. Thus the nonzero net current through the system is also an important property to distinguish the non-equilibrium system from the equilibrium condition, where the detailed balance is violated. The quantitative study of non-equilibrium systems, especially the one far away from equilibrium & steady state requires calculations of dynamical observable [69] such as changing rate of state function, which relies on *Large Deviation theorem (LDT)* from advanced probability theory [85]. In short, LDT estimates the probability that the empirical average approaches a value other than the real average. This deviation could be regarded as rare events. Here we defer this action to our future work, but present a qualitative prove of the presence of the non-equilibrium condition by using typical tool, *Fluctuation-Dissipation theorem*.

4.8.2 Verify the non-equilibrium system by FDT

In above section 4.7.2, our theory absorbed the dynamics of size change into the displacement by using perturbation theory, the resulted coupled dynamics offers an evidence of non-equilibrium condition by the appearance of "memory kernel". Here we use the typical tool, the Fluctuation-Dissipation Theorem (FDT), to quantitatively prove the presence of non-equilibrium:

$$S_x(\omega) = \frac{2k_BT}{\omega}\chi(\omega) \tag{4.8.1}$$

here x is a observable or signal; S is the Fourier transform of signal into frequency space; χ is the response function or susceptibility, which usually measured under a oscillatory field. For a proof, see [86]. Here we only give a brief explanation: in general, FDT states that the autocorrelation function which is a measure of the intensity of spontaneous fluctuations is proportional to the imaginary part of the response function, which quantifies the energy dissipation by the environment due to an external perturbation.

This tool has proved to be a successful way to check the deviation from equilibrium.

In practical terms, it is a complicated method. First, this method requires two separate measurements including the correlation function and the response function be available. Second, the measurement of the response function depends on a measurement of the effect of a weak external force. Third, it is hard to guarantee the measurement are made within the linear response regime, that is, with a small perturbation. Below we use a AC force with a finite magnitude as small perturbations.

We exerted an periodical AC force: $f(t) = Msin(\omega t)$ as an extra term of the equation of motion of the dropet. Then the displacement would emerges a periodical movement as expected and we can guess it can be expressed by a simple Fourier expansion $r_1 cos\omega t + r_2 sin\omega t) sin\omega t$.

Then the out phase component of the driven displacement are calculated by:

$$\int_0^{T_{tot}} r(t) sin\omega t dt = \sum_{n=0}^{n=n_{max}} r(n\Delta t) sin(\omega n\Delta t)\Delta t \tag{4.8.2}$$

Here, n and n_{max} are the number of steps that should be converted to a real number by multiplying with Δt; All raw data below are without the normalizing prefactor $1/T_{tot}$ and the time step Δt. Also for the integration in equation 80, the target is to get the coefficient of the Fourier series, where in the integral for the guessed displacement:

$$\int_0^{T_{tot}} (r_1 cos\omega t + r_2 sin\omega t) sin\omega t dt \simeq \frac{1}{2} r_2 T_{tot} \tag{4.8.3}$$

Thus one should multiply the raw integration with 2 to get the Fourier component r_2. The resulting out-of-phase component is measured for different magnitudes of the AC force. Below we take $\omega = 5$ and simulated the intermediate case $\alpha = 3$ as an example. The measured slope is the response. Again as we mentioned earlier, one should convert it with several prefactors.

Then by assuming the driven system is a damped harmonic oscillator, one can get the analytical expression of the susceptibility (response).

$$Re\chi = \frac{k}{k^2 + \gamma^2\omega^2}; Im\chi = \frac{-\gamma\omega}{k^2 + \gamma^2\omega^2}$$

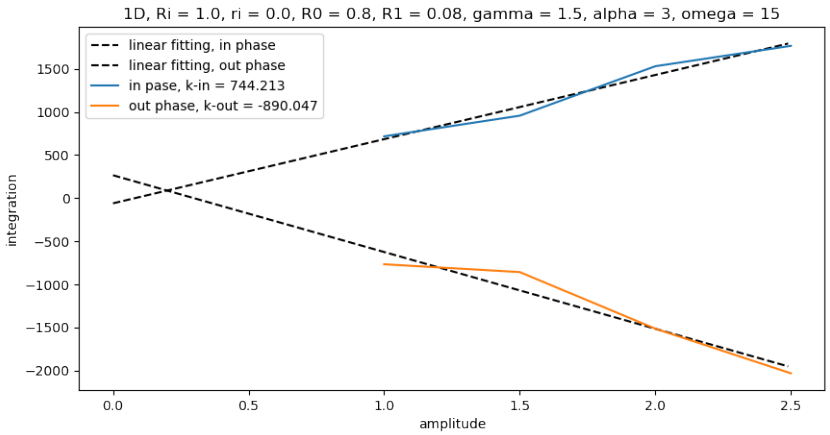
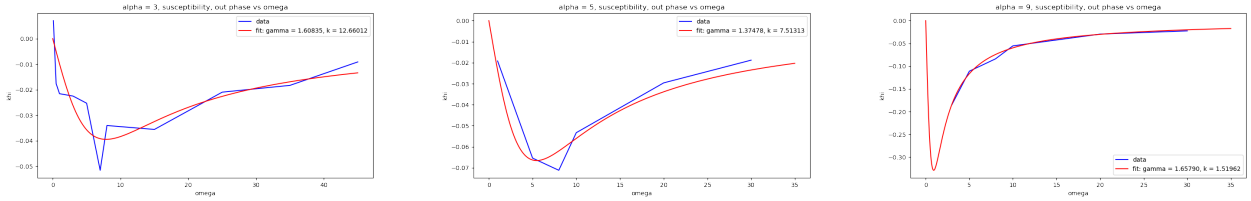


Figure 4.8.1: Measure response from raw data of Fourier component v.s amplitude



(a) $\alpha = 3$ (b) $\alpha = 5$ (c) $\alpha = 9$

Figure 4.8.2: Fitting $\Im\chi$

(4.8.4)

The physical meaning of the real part is the effective spring constant and the imaginary part is the dissipation due to the external force.

We want to fit the numerical results above in this form by regarding the γ and k as undecided constant and ω as a variable.

By converting the raw data to the real magnitude of Fourier component as mentioned at the beginning of this section, the linear fitted parameter has a good agreement with the one we really use, in which the friction coefficient $\gamma = 1.5$; effective spring depending on saturated radius $D\bar{R}^2$.

Then we turn to the measurement of autocorrelation function to check the relation be-

tween imaginary component of χ and Fourier transformed auto-correlation function for the validation of FDT:

$$c(\omega) = \frac{2k_BT/\gamma}{\omega^2 + k^2/\gamma^2} = \frac{2k_BT}{\omega} Im\chi(\omega) \tag{4.8.5}$$

The $c(\omega)$ is calculated by pick the real part of the discrete Fourier transform (to consistent with equation 80). Noticed in equation 80, we use ordinary frequency ω, here we first use the angular frequency k which be generally used in built-in functions of Python packages:

$$\Omega_k = Re \sum_{n=-n_{max}}^{n=n_{max}} c(n\Delta t) cos(\frac{2\pi}{T_{tot}} kn\Delta t)\Delta t \tag{4.8.6}$$

by relation $T_{tot} = n_{max}\Delta t$, we can rewrite equation 5 as:

$$\sum_{n=-n_{max}}^{n=n_{max}} c(n\Delta t) cos(\frac{2\pi}{n_{max}} kn)\Delta t \tag{4.8.7}$$

For an ACF with noisy signal in larger lag times, we could fitted the discrete signal $c(n\delta t)$ as an analytic polynomial function $c(t)$. This continuous signal can be transferred by continuous time Fourier transform (CTFT):

$$\Omega_\xi = \int_{-T_{tot}}^{T_{tot}} c(t) e^{-2\pi i t\xi} dt \tag{4.8.8}$$

For the angular frequency ξ in CTFT, we should translate it into k in DTFT. The exponential argument $t\xi$ in CTFT can be analogy as a discrete expression $\xi\Delta tn$, which should equals to the argument in equation 4.8.7: $k\frac{n}{n_{max}} = \xi\Delta tn$.

We will follow this with first a checking of simple case FT, e^{-ax^2}:

One can see the CTFT and DTFT have a perfect agreement with theory (figure 4.8.3).

Finally we apply this routine to the fitting of ACF signal. Noticed to consistent with the ordinary frequency ω used in simulation as shown in equation 4.8.2, one should pick the k^{th} discrete Fourier component (equation 4.8.6 & 4.8.7), for which $\frac{2\pi}{n_{max}} k = \omega\Delta t$. Then, with a consistent ω one need to times Δt to convert to the real value, as in the built-in function, there is a default time step equals one, which is not the case here. Finally, again, one need to convert all frequency to a standard dimension. Here we use ξ mentioned above:

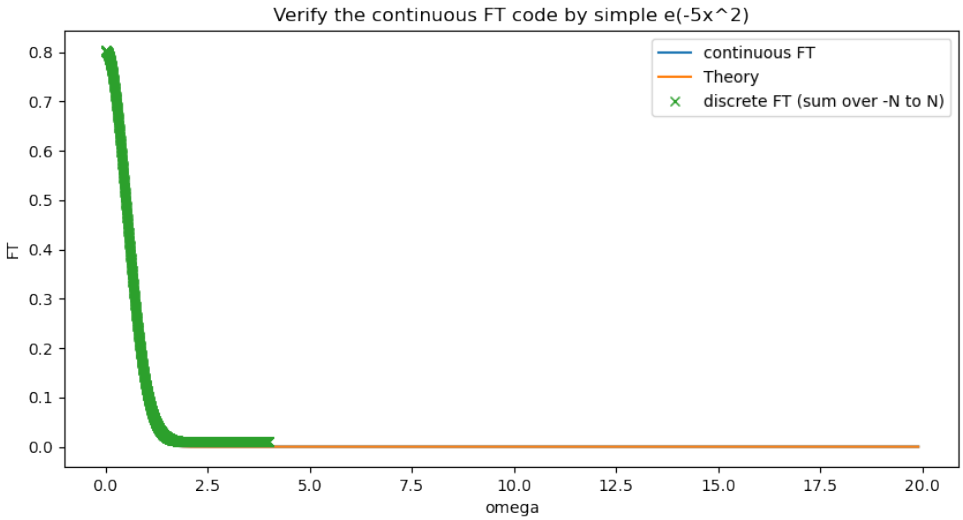
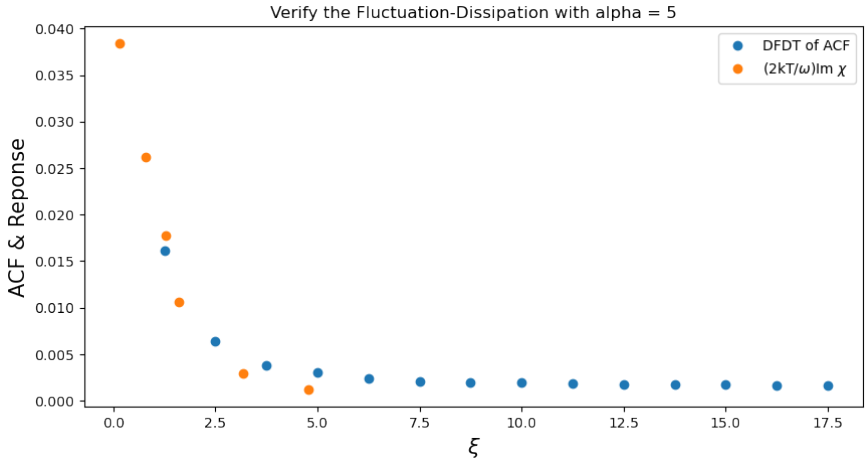


Figure 4.8.3: Verify the DTFT and CTFT by e^{-5x^2}

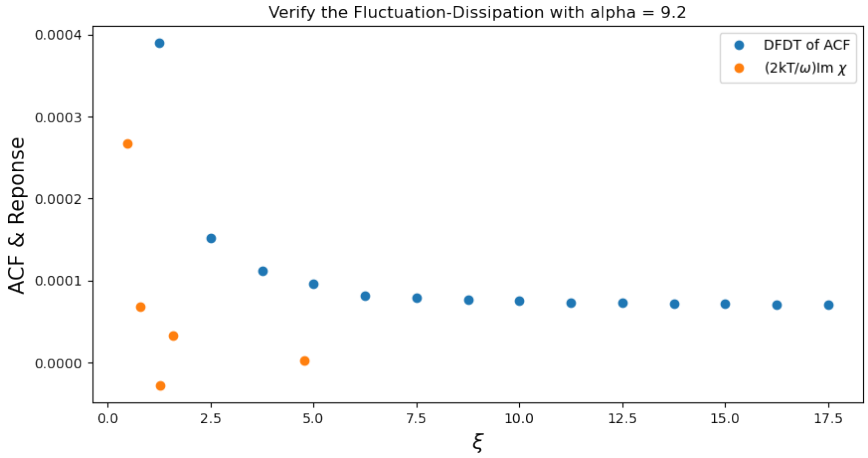
$$2\pi k\frac{n}{n_{max}} = 2\pi\xi\Delta tn = \omega n\Delta t \tag{4.8.9}$$

In figure 4.8.4 are two set of curves with $\alpha = 5$ and 9.2, we can easily see at small α the equilibrium condition still holds, but with a larger one, the FDT is violated. In the later case, the deviations are significant at low frequency region, which qualitatively agree with Large Deviation theorem that those events should be rare, and its frequencies are expected to be low.

In summary we have shown for a simple model of arrested phase-separation by an external potential that the resulting droplets constantly consume free energy and are in a non-equilibrium state.



(a) $\alpha = 5$



(b) $\alpha = 9.2$

Figure 4.8.4: Verify the FDT by DTFT, the blue dots are Fourier transformed Autocorrelation function of displacement r; the orange dots are imaginary part of response function χ from sequence of perturbed simulations

APPENDIX A

Numerical integration of the Langevin equation in dimensionless units

We start from a general one-dimensional Langevin equation:

$$m\ddot{x} + \gamma\dot{x} = F(x) + \eta(t) \tag{A.0.1}$$

Here η is the white noise with zero mean and with an autocorrelation function $< \eta(t)\eta(t') >= 2k_BT\delta(t-t')$. We next discretize this equation by integrating over a time interval Δt with

$$t_n = n\Delta t; x_n = x(t_n); v_n = v(t_n) \tag{A.0.2}$$

This gives:

$$\int_{t_n}^{t_n+\Delta t} \frac{d^2x}{dt^2}dt + \int_{t_n}^{t_n+\Delta t} \gamma\frac{dx}{dt}dt = \int_{t_n}^{t_n+\Delta t} F(x)dt + \int_{t_n}^{t_n+\Delta t} \eta(t)dt \tag{A.0.3}$$

This is approximated as

$$m(v_{n+1} - v_n) + \gamma(x_{n+1} - x_n) = \int_{t_n}^{t_n+\Delta t} F(x)dt + \int_{t_n}^{t_n+\Delta t} \eta(t)dt \tag{A.0.4}$$

at time t_{n+1} the force is expanded in a Taylor expansion, $F(x_n)+F'(x_n)v_n(t-t_n)+...\mathcal{O}(\Delta t^2)$. Inserting this into the integral we get:

$$\int_{t_n}^{t_n+\Delta t} Fdt = F(x_n)\Delta t + \frac{1}{2}F'(x_n)v_n\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{A.0.5}$$

Since $\eta(t)$ & $\eta(t')$ are uncorrelated random variables, the integral of the white noise must be a Gaussian random variable with zero expectation. Its variance can be calculated as:

$$\sigma_n^2 = \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t_{n+1}} dt'' < \eta(t')\eta(t'') >= 2k_B T\gamma\Delta t \tag{A.0.6}$$

thus the magnitude of the integral of the noise is $2k_BT\gamma\Delta t$, we here name it as Δn. This expression indicates that the random force will dominant when time step is small enough. With preparing all integral up to a range within time step, now the Langevin equation can be translated to a update rule:

$$x_{n+1} = x_n + v_n\Delta t + \mathcal{O}(\Delta t^2); v_{n+1} = v_n - \gamma\frac{v_n\Delta t}{m} - \frac{F(x_n)\Delta t}{m} + \frac{\Delta n}{m} + \mathcal{O}(\Delta t^2) \tag{A.0.7}$$

We now convert all quantities into a dimensionless manner. The important step is define a characteristic time: $\tau_0 = \frac{m}{\tau}$. This selection guaranteed when $\Delta t < \tau_0$, the inertial effect would dominant, on the contrary the viscous effect will play a main role. Then the dimensionless time step should be $\bar{\Delta t} = \frac{\Delta t}{\tau_0}$.

Then we define the characteristic energy by equipartition theorem: $\frac{1}{2}m < v^2 >= \frac{1}{2}k_BT$. Thus the characteristic velocity and displacement can be derived as:

$$v_0 = \sqrt{\frac{k_BT}{m}}; x_0 = v_0\tau_0 \tag{A.0.8}$$

similarly the dimensionless quantities can be obtained by dividing the characteristic one from raw value. This makes the updating rule in A.0.7 can be rewrite as follows:

$$\bar{x_{n+1}} = \bar{x_n} + \frac{v_n\Delta t}{x_0} = \bar{x_n} + \frac{v_n\Delta t}{v_0\tau_0} = \bar{x_n} + \bar{v_n}\Delta t \tag{A.0.9}$$

$$\bar{v_{n+1}} = \bar{v_n} - \gamma\bar{v_n}\frac{\Delta t}{m}\tau_0 - \frac{F(x_n)\bar{\Delta t}\tau_0}{mv_0} + \frac{\Delta n}{mv_0} = \bar{v_n} - \bar{v_n}\bar{\Delta t} - \frac{F(x_n)\bar{\Delta t}\tau_0}{mv_0} + \frac{\Delta n}{mv_0} \tag{A.0.10}$$

the reorganization of the second term is hold by $\frac{\gamma\tau_0}{m} = 1$ which is the definition of τ_0. Then comparing the corresponding terms, we can further define characteristic force and random noise:

$$F_0 = \frac{mv_0}{\tau_0} = \frac{mv_0^2}{\tau_0 v_0} = \frac{k_B T}{x_0}; \bar{\Delta n} = \frac{\Delta n}{mv_0} \tag{A.0.11}$$

the variance, that is the magnitude of the noise can be estimated as:

$$\bar{\sigma^2} = \frac{2k_B T\gamma\Delta t}{m^2 v_0^2} = \frac{2k_B T\Delta t}{mv_0^2\tau_0} = 2\bar{\Delta} t \tag{A.0.12}$$

this lead to the expression in 3.4.3.

APPENDIX B

Detailed truncation of the spherical harmonics in the surface diffusion model

3.6.3 is an inhomogeneous linear equation in the coefficients C_n. The dependence on t has been left out, as we know that the solutions will be time-independent. If we multiply by $\cos(m\pi x/L)$ with $m \neq 0$, then the equation becomes homogeneous. In general it is

$$-\alpha \sum_{n=0}^{\infty} C_n \int_0^{L/2} \cos(n\pi x/L)\cos(m\pi x/L)\,dx - D\frac{\pi^2}{2L}C_m = 0 \tag{B.0.1}$$

The integral in (B.0.1) is readily carried out, and we end up with a set of homogeneous linear equations.

Then, in the steady state, we can replace the coefficients in equation 3.6.5 by their steady state values, which we will denote as $C_l^{(0)}$. The equations for those coefficients, obtained by multiplying the steady state equation by the $f_l(x)$'s and integrating are

$$0 = S\delta_{l,0} - (Dl(l+1)+\epsilon)C_l^{(0)} - \alpha \sum_{l'} C_{l'}^{(0)} \int_{-1}^{x_0} f_l(x) f_{l'}(x)\,dx \tag{B.0.2}$$

We assume that the reactive surface extends from $x = -1$ to $x = x_0$. In the calculations of current interest to us, we set $x_0 = 0$.

APPENDIX C

Calculation of the interaction energy of a rigid disk

We continue from 4.3.3. We first do the θ integral:

$$\int_0^{2\pi} d\theta e^{-\frac{xrcos\theta}{\sigma^2}} = 2\pi I_0(\frac{xr}{\sigma^2}) \tag{C.0.1}$$

which is the 0^{th} modified Bessel function of the 1^{st} kind. Thus the interaction energy can be rewrite as:

$$U_{int} = U_0 e^{-\frac{x^2}{2\sigma^2}} 2\pi \int_0^R r dr I_0(\frac{xr}{\sigma^2}) e^{-\frac{r^2}{2\sigma^2}} \tag{C.0.2}$$

Now we define dimensionless variables: $\bar{x} = \frac{x}{\sigma}; \bar{r} = \frac{r}{\sigma}$ and rewrite the integral as:

$$U_{int} = 2\pi U_0 e^{-\frac{\bar{x}^2}{2\sigma^2}} \sigma^2 \int_0^R \bar{r} d\bar{r} I_0(\frac{\bar{x}\bar{r}}{\sigma^2}) e^{-\frac{\bar{r}^2}{2}} \tag{C.0.3}$$

then calculate the force by chain rule: $F = -\frac{dU_{int}}{dx} = -\frac{dU_{int}}{\sigma d\bar{x}}$ then:

$$|F| = 2\pi U_0 \sigma | - \bar{x} e^{-\frac{\bar{x}^2}{2}} I + e^{-\frac{\bar{x}^2}{2}} \int_0^{\frac{R}{\sigma}} \bar{r} d\bar{r} I_0 e^{-\frac{\bar{r}^2}{2}} | \tag{C.0.4}$$

then use recursion relation $\frac{dI_0(x)}{ddx} = I_1(x)$, above equation becomes:

$$2\pi U_0 \sigma | \int_0^{\frac{R}{\sigma}} \bar{\sigma}^2 d\bar{r} I_1 e^{-\frac{\bar{r}^2}{2}} + e^{-\frac{\bar{x}^2}{2}} \int_0^{\frac{R}{\sigma}} \bar{r} d\bar{r} I_0 e^{-\frac{\bar{r}^2}{2}} | e^{-\frac{x^2}{2}} \tag{C.0.5}$$

by defining $B(\bar{x}, \bar{r}) = \bar{r} I_1(\bar{x}, \bar{r}) - \bar{x} I_0(\bar{x}, \bar{r})$, we finally clean the look:

$$f(\bar{x}, \bar{R}) = e^{-\frac{\bar{x}^2}{2}} | \int_0^{\frac{R}{\sigma}} d\bar{r} \bar{r} e^{-\frac{\bar{r}^2}{2}} B(\bar{x}, \bar{r}) | \tag{C.0.6}$$

REFERENCES

[1] Eric Vanden-Eijnden et al. Transition-path theory and path-finding algorithms for the study of rare events. *Annual review of physical chemistry*, 61:391–420, 2010.

[2] QH Zeng, AB Yu, and GQ Lu. Multiscale modeling and simulation of polymer nanocomposites. *Progress in polymer science*, 33(2):191–269, 2008.

[3] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-grained protein models and their applications. *Chemical reviews*, 116(14):7898–7936, 2016.

[4] Martin Karplus. Development of multiscale models for complex chemical systems: from h+ h2 to biomolecules (nobel lecture). *Angewandte Chemie International Edition*, 53(38):9992–10005, 2014.

[5] Michael P Allen, Daan Frenkel, and J Talbot. Molecular dynamics simulation using hard particles. *Computer physics reports*, 9(6):301–353, 1989.

[6] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.

[7] Kenno Vanommeslaeghe, Elizabeth Hatcher, Chayan Acharya, Sibsankar Kundu, Shijun Zhong, Jihyun Shim, Eva Darian, Olgun Guvench, P Lopes, Igor Vorobyov, et al. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–690, 2010.

[8] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[9] Daan Frenkel, Berend Smit, and Mark A Ratner. *Understanding molecular simulation: from algorithms to applications*, volume 2. Academic press San Diego, 1996.

[10] Brandon Malone, James Chen, Qi Wang, Eliza Llewellyn, Young Joo Choi, Paul Dominic B Olinares, Xinyun Cao, Carolina Hernandez, Edward T Eng, Brian T Chait, et al. Structural basis for backtracking by the sars-cov-2 replication–transcription complex. *Proceedings of the National Academy of Sciences*, 118(19), 2021.

[11] Xiaocheng Shang and Martin Kroger. Time correlation functions of equilibrium and nonequilibrium langevin dynamics: Derivations and numerics using random numbers. *Siam Review*, 62(4):901–935, 2020.

[12] Yuan-cheng Fung. A first course in continuum mechanics. *Englewood Cliffs*, 1977.

[13] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):826–843, 2011.

[14] David E Shaw, Martin M Deneroff, Ron O Dror, Jeffrey S Kuskin, Richard H Larson, John K Salmon, Cliff Young, Brannon Batson, Kevin J Bowers, Jack C Chao, et al. Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*, 51(7):91–97, 2008.

[15] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2):141–151, 1999.

[16] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

[17] Brooke E Husic and Vijay S Pande. Markov state models: From an art to a science. *Journal of the American Chemical Society*, 140(7):2386–2396, 2018.

[18] Alessandro Laio and Francesco L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12):126601, 2008.

[19] Omar Valsson and Michele Parrinello. Variational approach to enhanced sampling and free energy calculations. *Physical review letters*, 113(9):090601, 2014.

[20] Ilias Bilionis and Phaedon-Stelios Koutsourelakis. Free energy computations by minimization of kullback–leibler divergence: An efficient adaptive biasing potential method for sparse representations. *Journal of Computational Physics*, 231(9):3849–3870, 2012.

[21] Eric Darve and Andrew Pohorille. Calculating free energies using average force. *The Journal of chemical physics*, 115(20):9169–9183, 2001.

[22] Sanghyun Park and Klaus Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *The Journal of chemical physics*, 120(13):5946–5961, 2004.

[23] Thomas Huber, Andrew E Torda, and Wilfred F Van Gunsteren. Local elevation: a method for improving the searching properties of molecular dynamics simulation. *Journal of computer-aided molecular design*, 8(6):695–708, 1994.

[24] Christophe Chipot and Andrew Pohorille. Free energy calculations. *Springer series in chemical physics*, 86:159–184, 2007.

[25] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[26] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of computational chemistry*, 13(8):1011–1021, 1992.

[27] Coffin et al. Retroviruses. In *Cold Spring Harbor symposia on quantitative biology*. Cold Spring Harbor Laboratory Press, 1997.

[28] Florian KM Schur, Wim JH Hagen, Michaela Rumlová, Tomáš Ruml, Barbara Müller, Hans-Georg Kräusslich, and John AG Briggs. Structure of the immature hiv-1 capsid in intact virus particles at 8.8 å resolution. *Nature*, 517(7535):505, 2015.

[29] James B Munro, Abhinav Nath, Michael Färber, Siddhartha AK Datta, Alan Rein, Elizabeth Rhoades, and Walther Mothes. A conformational transition observed in single hiv-1 gag molecules during in vitro assembly of virus-like particles. *Journal of virology*, 88(6):3577–3585, 2014.

[30] Chinh Tran-To Su, Chee-Keong Kwoh, Chandra Shekhar Verma, and Samuel Ken-En Gan. Modeling the full length hiv-1 gag polyprotein reveals the role of its p6 subunit in viral maturation and the effect of non-cleavage site mutations in protease drug resistance. *Journal of Biomolecular Structure and Dynamics*, pages 1–12, 2017.

[31] Enrico Guarnera and Igor N Berezovsky. Allosteric sites: remote control in regulation of protein activity. *Current opinion in structural biology*, 37:1–8, 2016.

[32] Joseph A Webb, Christopher P Jones, Leslie J Parent, Ioulia Rouzina, and Karin Musier-Forsyth. Distinct binding interactions of hiv-1 gag to psi and non-psi rnas: implications for viral genomic rna packaging. *RNA*, 19(8):1078–1088, 2013.

[33] Mauricio Comas-Garcia, Siddhartha AK Datta, Laura Baker, Rajat Varma, Prabhakar R Gudla, and Alan Rein. Dissection of specific binding of hiv-1 gag to the'packaging signal'in viral rna. *Elife*, 6, 2017.

[34] Robert Dick, Donna Mallery, Volker Vogt, and Leo James. Ip6 regulation of hiv capsid assembly, stability, and uncoating. *Viruses*, 10(11):640, 2018.

[35] HPG Drewry and NA Seaton. Continuum random walk simulations of diffusion and reaction in catalyst particles. *AIChE Journal*, 41(4):880–893, 1995.

[36] Howard C Berg and Edward M Purcell. Physics of chemoreception. *Biophysical journal*, 20(2):193–219, 1977.

[37] M von Smoluchowski. Drei vortrage uber diffusion, brownsche bewegung und koagulation von kolloidteilchen. *Zeitschrift fur Physik*, 17:557–585, 1916.

[38] Frank C Collins and George E Kimball. Diffusion-controlled reaction rates. *Journal of colloid science*, 4(4):425–437, 1949.

[39] O Bénichou, D Grebenkov, P Levitz, C Loverdo, and R Voituriez. Optimal reaction time for surface-mediated diffusion. *Physical review letters*, 105(15):150606, 2010.

[40] J-F Rupprecht, O Bénichou, DS Grebenkov, and R Voituriez. Kinetics of active surface-mediated diffusion in spherically symmetric domains. *Journal of Statistical Physics*, 147(5):891–918, 2012.

[41] Le Yang, David Sept, and AE Carlsson. First-contact time to a patch in a multidimensional potential well. *Physical Review E*, 76(2):021911, 2007.

[42] Changsun Eun. Effect of surface curvature on diffusion-limited reactions on a curved surface. *The Journal of chemical physics*, 147(18):184112, 2017.

[43] Kuo Chen Chou and Guo Ping Zhou. Role of the protein outside active site on the diffusion-controlled reaction of enzymes. *Journal of the American Chemical Society*, 104(5):1409–1413, 1982.

[44] Claire E Plunkett and Sean D Lawley. Bimolecular binding rates for pairs of spherical molecules with small binding sites. *Multiscale Modeling & Simulation*, 19(1):148–183, 2021.

[45] Denis S Grebenkov. First passage times for multiple particles with reversible target-binding kinetics. *The Journal of chemical physics*, 147(13):134112, 2017.

[46] Denis S Grebenkov. Reversible reactions controlled by surface diffusion on a sphere. *The Journal of chemical physics*, 151(15):154103, 2019.

[47] Yurii A Makhnovskii, Alexander M Berezhkovskii, Sheh-Yi Sheu, Dah-Yen Yang, Jimmy Kuo, and Sheng Hsien Lin. Stochastic gating influence on the kinetics of diffusion-limited reactions. *The Journal of chemical physics*, 108(3):971–983, 1998.

[48] Oleg V Bychuk and Ben O'Shaughnessy. Anomalous diffusion at liquid surfaces. *Physical review letters*, 74(10):1795, 1995.

[49] J-F Rupprecht, O Bénichou, DS Grebenkov, and R Voituriez. Exact mean exit time for surface-mediated diffusion. *Physical Review E*, 86(4):041135, 2012.

[50] Denis S Grebenkov. Paradigm shift in diffusion-mediated surface phenomena. *Physical Review Letters*, 125(7):078102, 2020.

[51] Albert Einstein. *Investigations on the Theory of the Brownian Movement*. Courier Corporation, 1956.

[52] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[53] Jean Perrin. *Brownian movement and molecular reality*. Courier Corporation, 2013.

[54] David E Blair. *Inversion theory and conformal mapping*, volume 9. American Mathematical Soc., 2000.

[55] Melissa Vellela and Hong Qian. Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the schlögl model revisited. *Journal of The Royal Society Interface*, 6(39):925–940, 2009.

[56] Doug McLean. *Understanding aerodynamics: arguing from the real physics*. John Wiley & Sons, 2012.

[57] E Weinan, Tiejun Li, and Eric Vanden-Eijnden. *Applied stochastic analysis*, volume 199. American Mathematical Soc., 2019.

[58] Vijay S Pande, Kyle Beauchamp, and Gregory R Bowman. Everything you wanted to know about markov state models but were afraid to ask. *Methods*, 52(1):99–105, 2010.

[59] Henry Eyring. The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3(2):107–115, 1935.

[60] Christoph Dellago, Peter Bolhuis, Phillip L Geissler, et al. Transition path sampling. *Advances in chemical physics*, 123(1), 2002.

[61] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *The Journal of chemical physics*, 125(8):084110, 2006.

[62] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Transition path theory for markov jump processes. *Multiscale Modeling & Simulation*, 7(3):1192–1219, 2009.

[63] E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66(5):052301, 2002.

[64] Han Cheng Lie and Jannes Quer. Some connections between importance sampling and enhanced sampling methods in molecular dynamics. *The Journal of chemical physics*, 147(19):194107, 2017.

[65] Michael Slutsky and Leonid A Mirny. Kinetics of protein-dna interaction: facilitated target location in sequence-dependent potential. *Biophysical journal*, 87(6):4021–4035, 2004.

[66] C Loverdo, O Bénichou, M Moreau, and R Voituriez. Enhanced reaction kinetics in biological cells. *Nature physics*, 4(2):134–137, 2008.

[67] PK Pollett. The generalized kolmogorov criterion. *Stochastic processes and their applications*, 33(1):29–44, 1989.

[68] U. Seifert. Configurations of fluid membranes and vesicles. *Advances in physics*, 46(1):13–137, 1997.

[69] Jürgen Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern physics*, 48(4):571, 1976.

[70] Fabian Knoch and Thomas Speck. Non-equilibrium markov state modeling of periodically driven biomolecules. *The Journal of chemical physics*, 150(5):054103, 2019.

[71] Steven Boeynaems, Simon Alberti, Nicolas L Fawzi, Tanja Mittag, Magdalini Polymenidou, Frederic Rousseau, Joost Schymkowitz, James Shorter, Benjamin Wolozin, Ludo Van Den Bosch, et al. Protein phase separation: a new phase in cell biology. *Trends in cell biology*, 28(6):420–435, 2018.

[72] David Zwicker, Markus Decker, Steffen Jaensch, Anthony A Hyman, and Frank Jülicher. Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. *Proceedings of the National Academy of Sciences*, 111(26):E2636–E2645, 2014.

[73] Joshua A Riback, Christopher D Katanski, Jamie L Kear-Scott, Evgeny V Pilipenko, Alexandra E Rojek, Tobin R Sosnick, and D Allan Drummond. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell*, 168(6):1028–1040, 2017.

[74] Avinash Patel, Hyun O Lee, Louise Jawerth, Shovamayee Maharana, Marcus Jahnel, Marco Y Hein, Stoyno Stoynov, Julia Mahamid, Shambaditya Saha, Titus M Franzmann, et al. A liquid-to-solid phase transition of the als protein fus accelerated by disease mutation. *Cell*, 162(5):1066–1077, 2015.

[75] Robert W Style, Tianqi Sai, Nicoló Fanelli, Mahdiye Ijavi, Katrina Smith-Mannschott, Qin Xu, Lawrence A Wilen, and Eric R Dufresne. Liquid-liquid phase separation in an elastic network. *Physical Review X*, 8(1):011028, 2018.

[76] Kathryn A Rosowski, Estefania Vidal-Henriquez, David Zwicker, Robert W Style, and Eric R Dufresne. Elastic stresses reverse ostwald ripening. *Soft Matter*, 16(25):5892–5897, 2020.

[77] Estefania Vidal-Henriquez and David Zwicker. Theory of droplet ripening in stiffness gradients. *Soft Matter*, 16(25):5898–5905, 2020.

[78] Xuefeng Wei, Jiajia Zhou, Yanting Wang, and Fanlong Meng. Modeling elastically mediated liquid-liquid phase separation. *Physical Review Letters*, 125(26):268001, 2020.

[79] Pierre Ronceray, Sheng Mao, Andrej Košmrlj, and Mikko P Haataja. Liquid demixing in elastic networks: cavitation, permeation, or size selection? *arXiv preprint arXiv:2102.02787*, 2021.

[80] Chuck Yeung, T Rogers, A Hernandez-Machado, and David Jasnow. Phase separation dynamics in driven diffusive systems. *Journal of statistical physics*, 66(3):1071–1088, 1992.

[81] Lorenzo Costigliola, David M Heyes, Thomas B Schrøder, and Jeppe C Dyre. Revisiting the stokes-einstein relation without a hydrodynamic diameter. *The Journal of chemical physics*, 150(2):021101, 2019.

[82] Frank Jülicher, Armand Ajdari, and Jacques Prost. Modeling molecular motors. *Reviews of Modern Physics*, 69(4):1269, 1997.

[83] Douglas E Smith, Sander J Tans, Steven B Smith, Shelley Grimes, Dwight L Anderson, and Carlos Bustamante. The bacteriophage φ29 portal motor can package dna against a large internal force. *Nature*, 413(6857):748–752, 2001.

[84] Michael E Cates and Julien Tailleur. Motility-induced phase separation. *Annu. Rev. Condens. Matter Phys.*, 6(1):219–244, 2015.

[85] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.

[86] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Course of theoretical physics*. Elsevier, 2013.