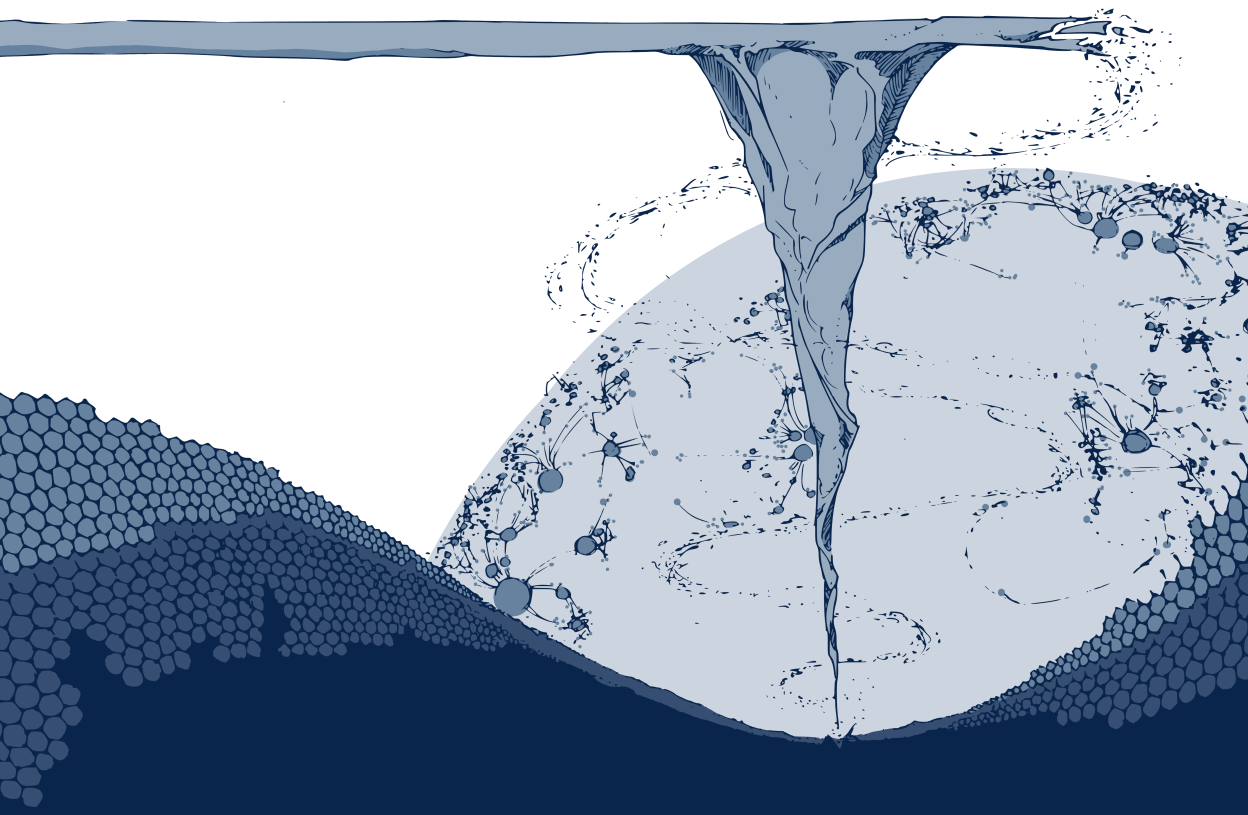
Stockholm University

Simple Models for Complex Nonequilibrium Problems in Nanoscale Friction and Network Dynamics

David Andersson



Doctoral Thesis in Theoretical Physics at Stockholm University, Sweden 2020

Simple Models for Complex Nonequilibrium Problems in Nanoscale Friction and Network Dynamics

David Andersson

Academic dissertation for the Degree of Doctor of Philosophy in Theoretical Physics at Stockholm University to be publicly defended on Friday 15 January 2021 at 14.00 in sal FB42, AlbaNova universitetscentrum, Roslagstullsbacken 21 and online via Zoom, public link is available at the department website.

Abstract

This doctoral thesis investigates three different topics: How friction evolves in atomically thin layered materials (2D materials); How social dynamics can be used to model grand scale common-pool resource games; Benchmarking of various image reconstruction algorithms in atomic force microscopy experiments. While these topics are diverse, they share being complex out-of-equilibrium systems. Furthermore, our approach to these topics will be the same: using simple models to obtain qualitative information about a system's dynamics. In the case of atomically thin layered materials, we will be expanding on the influential Prandtl-Tomlinson model and obtain an improved model constituting a substantial improvement in the theoretical description of friction in these systems. In the context of social dynamics, we will introduce a novel model representing a new approach to consensus rates on social networks in relation to society spanning coordination problems. For the image reconstruction project, our ambition is to investigate a new method for recreating free-energy surfaces based on AFM experiment, however, for this project only preliminary results are included.

Keywords: *tribology, nanofriction, 2d materials, graphene, image reconstruction, social dynamics, common-pool resource, collective action, simple models.*

Stockholm 2020
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-186930

ISBN 978-91-7911-378-0
ISBN 978-91-7911-379-7

Department of Physics

Stockholm University, 106 91 Stockholm

SIMPLE MODELS FOR COMPLEX NONEQUILIBRIUM PROBLEMS IN NANOSCALE FRICTION AND NETWORK DYNAMICS

David Andersson

Stockholm
University

Simple Models for Complex Nonequilibrium Problems in Nanoscale Friction and Network Dynamics

David Andersson

ISBN print 978-91-7911-378-0
ISBN PDF 978-91-7911-379-7

Cover art by Alexandra Polyakova (@polykalex)

Printed in Sweden by Universitetsservice US-AB, Stockholm 2020

To my mother,
my northern star,
Lena Wiklund.

Abstract

This doctoral thesis investigates three different topics: how friction evolves in atomically thin layered materials (2D materials); how social dynamics can be used to model large scale common-pool resource games; benchmarking of various image reconstruction algorithms in atomic force microscopy experiments. While these topics are diverse, they all are complex out-of-equilibrium systems. Furthermore, our approach to these topics will be the same: using simple models to obtain qualitative information about the dynamics. In the case of atomically thin layered materials, we will be expanding on the influential Prandtl-Tomlinson model and obtain a substantial improvement in the theoretical description of friction in these systems. In the context of social dynamics, we will introduce a novel model representing a new approach to consensus rates in social networks in relation to society spanning coordination problems. For the image reconstruction project, our ambition is to investigate a new method for recreating free-energy surfaces based on atomic force microscopy experiments. However, for this project only preliminary results are included.

Sammanfattning

Den här doktorsavhandlingen behandlar tre olika områden: hur friktion ter sig i tvådimensionella material, hur socialdynamik kan användas för att modellera storskaliga samarbetsutmaningar kring gemensamma vitala resurser samt rankning av olika procedurer för att återskapa ytor från data genererad med atomkraftsmikroskåp. Dessa områden kan förefalla ha lite gemensamt, men de är alla komplexa icke-jämviktssystem. Dessutom kommer de att behandlas med samma verktyg: så kallade förenklade modeller kommer att användas för att erhålla kvalitativ information om dynamiken. I kontexten tvådimensionella material kommer vi att expandera Prandtl-Tomlinsonmodellen och signifikant förbättra den teoretiska beskrivningen av friktion i sådana system. Inom området socialdynamik kommer vi att introducera en modell som på ett nytt sätt relaterar hur grupperingar och konsensus uppstår i stora gemensamma samhällsutmaningar. Till slut kommer vi att presentera preliminära resultat kring hur ytor kan återskapas från atomkraftsmikroskåp, detta projekt är emellertid inte slutfört i skrivande stund.

List of accompanying papers

Paper I – D. Andersson, A. S. de Wijn. "Understanding the friction of atomically thin, layered materials". *Nature Communications*, vol. 11, no. 1, pp. 1–7, 2020 [1].

My contribution: The model itself was developed through successive iterations across multiple interactions between me and my supervisor. While my supervisor had the initial idea of expanding the Prandtl-Tomlinson model, I made significant contributions to its development. I also did a literature search to identify empirical data that me and my supervisor successfully related to our model. Furthermore I wrote and ran all the simulations from scratch, and made all the figures. The paper was co-written with my supervisor.

Paper II – J. Roadnight Sheehan, D. Andersson, A. S. de Wijn. "Thermal effects and spontaneous frictional relaxation in atomically thin layered materials". *arXiv*, id. 2012.00371, 2020 [2]. Submitted to *Physical Review B*.

My contribution: I did all the coding for the simulations, including various algorithms for data manipulation. I also subsequently generated all the numerical data for the project. I was an active part in discussing how to interpret the simulation results and relate them to experiment, however, the analytical calculations of rates were done by my co-authors. The paper was written collaboratively. I made figures: 1-6(a).

Paper III – D. Andersson, S. Bratsberg, A. K. Ringsmuth, A. S. de Wijn. "Dynamics of collective action to conserve a large common-pool resource". *arXiv*, id. 2012.00892, 2020 [3]. Submitted to *Nature Human Behavior*.

My contribution: I formulated the basic idea of using an agent based network approach, as well as a rudimentary first version of the model. I then co-supervised a master who produced much of the base code for the simulations. After the student finished however, I made significant additions and alterations to the code. The model was continuously developed through regular discussions with my co-authors during the early stages of the project, and I had a prominent role in this. I have contributed all the data and every figure to the paper. The paper was co-written with my co-authors.

Reused material from the licentiate thesis

This PhD thesis is the continuation of the licentiate thesis called "On the Evolving Friction of Layered Materials and the Prospect of Their Image Reconstruction", unpublished and written by the same author in 2019. That thesis partially covered the same materials as chapters: 1,2 and 6 in this thesis. The results from chapters 1 and 2 are identical since they were then already published in Paper I, but the chapters have been modified in terms of presentation. Chapter 6 has both new material, and new presentation, but there is some overlap with the previous thesis.

Contents

List of Abbreviations, Concepts, and Conventions

Here we list some important abbreviations, concepts, conventions, definitions and notation introduced in this thesis. N.B. In many cases a term might have multiple, or a more general meaning than the one given here, the present statement then reflects how the term is used within this thesis in particular. Priority has also been given to write short and accessible explanations. We refer to the accompanied page reference for the proper definition. We hope that it will serve as a reference for the reader. The items appear in alphabetical order.

AFM

Atomic Force Microscope, an instrument consisting of a tip, a cantilever and a support used to study microscopic friction (ref. page 9).

Agent based model and agent interaction

Autonomous agents in a network that interact with adjacent agents in some prescribed way. In this work they strive to convince other agents of their own opinions (ref. page 75).

Clustered scale-free network

A network where the shortest distance between any two nodes is small, and moreover nodes tend to gather into communities (ref. page 63).

Clustering, communities, and modularity

A community is a set of topologically connected nodes, that is characterized by high modularity. Clusters are nets of nodes sharing a state, they can transcend community borders (ref. page 66).

Collective action and the tragedy of the commons

The coordinated response needed to address complex system-wide challenges (collective action (ref. page 60)), in order to avoid collapse due to actors acting in apparent self interest (tragedy of the commons (ref. page 58)) .

(In-)Commensurate

Two periodic surfaces are said to be commensurate if their periodicity matches up to some degree. If they do not, then they are incommensurate (ref. page 23).

Common pool-resource

A limited and vital resource that is shared by a community and is openly available to all members. The community must limit their consumption in order not to deplete it. Members are said to be cooperators or defectors depending on how they comply with this (ref. page 60).

Decay path/chain

The sequence of potential minima through which an AFM tip relaxes (ref. page 48).

External field (social dynamics)

A global parameter that incentivizes all agents in a network to become more cooperative or defective (ref. page 76).

Friction layer dependence

The observed friction decreases when multiple layers of some atomically thin material is stacked (ref. page 34).

Friction strengthening

The phenomenon when friction initially tends to increase in friction experiments on thin sheets of layered materials (ref. page 16).

Influencer

A highly connected node in a social network, this gives a strong position to spread their opinions (ref. page 88).

The Jarzynski equality

A fundamental relationship between average work and free energy of a thermodynamic system (ref. page 99).

Kramers and decay rates

The rate at which an AFM tip leaves a potential minimum due to thermal kicks is referred to as the decay rate. This can be modeled by the Kramers rate (ref. page 41).

Langevin dynamics

We use this as a protocol used to incorporate thermal fluctuations into molecular dynamics (MD) simulations (ref. page 14).

(Instantaneous) Lateral force

The momentaneous force on the AFM tip in the direction opposite to sliding. The friction force is the time average of the lateral force (ref. page 10).

Layer distortion (q)

This is a degree of freedom introduced by us to capturing how a 2D sheet distorts as it interacts with a substrate and a tip (ref. page 17).

Layered material

A material made up of potentially several layers of some material, we will typically require that the materials permit mono-layers, also called 2D materials (ref. page 15).

Lucy Richardson deconvolution LRD

The Lucy-Richardson Deconvolution (algorithm), a scheme for reconstructing an image from corrupted data (ref. page 107).

Maximum allowed sheet distortion ($q_{\max}$)

The maximum allowed deformation of the sheet in a substrate-sheet-AFM system. (ref. page 27)

Modified PT-model

The model proposed in this thesis to extend the PT-model to layered materials (ref. page 20).

Moiré pattern and the lattice parameter ratio (γ)

The pattern that emerges as two periodic surfaces with non-matching lattice parameters are stacked, due to them periodically being in and out of phase. The mismatch can be characterized by the ratio of the lattice parameters, γ (ref. page 30).

Point spread function

A function that captures how an image is corrupted by data corrupted by a Poisson distributed noise (ref. page 105).

PT-model

The Prandtl-Tomlinson model – a common model for the friction in atomic systems (ref. page 7)

Potential landscape

The two dimensional potential energy surface of the sheet-substrate superposition in the modified PT model (ref. page 26).

Sheet

Several stacked layers of some 2D material, see Layered material.

Social and opinion dynamics

How social norms, opinions, and behaviours spread between people. In particular how they transition from individual to group scale (ref. page 60).

Stick-slip motion

A motion exhibited by a tip in the PT model*. The tip will periodically be stuck climbing the potential and slipping over barriers (ref. page 8).

*Stick-slip motion is a quite general phenomenon and is found in microscopic as well as macroscopic systems, but this reference suffices for our purposes.

Supported and suspended sheets

Supported sheets are deposited onto some substrate, whereas suspended sheets do not have anything underneath them (ref. page 22).

Thermal/spontaneous relaxation

The spontaneous decay when an AFM tip and sheet relaxes towards a potential minimum as it is stopped (ref. page 46).

Thermally activated slips

When an AFM tip slides in the presence of thermal fluctuations, slips can occur sooner than anticipated due to random kicks to the tip. This can be seen as a kind of lubrication which is referred to as thermolubricity (ref. page 11).

Thermolubricity

See thermally activated slips.

Transition state

A temporary state which a system transitions through when it goes between two metastable states (ref. page 43).

WHAM

The Weighted Histogram Analysis Method, a way of calculating the free energy surface with respect to some reaction coordinate (ref. page 103).

An Introduction

This PhD thesis is the result of four years full time graduate studies. The project was jointly supported by Stockholm University and the Norwegian University of Science and Technology. The thesis is a compilation thesis, meaning that it consists of a set of papers which resulted from the research conducted during the project, as well as a comprehensive introduction and summary to these papers. The papers follow a thematic order, which incidentally corresponds well to the order of time spent on each theme. In the first part of the thesis we will be investigating nanofriction and how simple models can be used to describe layered materials within this context. We will then spend some time working out the details of how these models behave in the presence of thermal noise. This covers Papers I and II. In the second part of the thesis we will, once again using simple models, investigate the social dynamics of large scale common-pool resource problems. This covers Paper III. Finally, in the concluding third part of the thesis we will be looking into image reconstruction schemes, and specifically how free energy surfaces can be reconstructed from repeated atomic force microscopy experiments. This is an ongoing project, and thus no finished manuscript will be supplied here. However, sufficient progress has been made that there are some preliminary results to report.

On a note of form, the organizational hierarchy of the thesis is as follows: Each topic will be analyzed in separate parts. Furthermore, each part is divided into two chapters, the first of which looks into background theory, and the second which presents the actual findings from the corresponding paper. The exception of this is Chapters 3, which contains all the information pertaining to Paper II. The third part is also an exception to this approach, since the availability of results there is still low. In terms of style, the papers are written with a scientific audience of experts in mind, and priority has accordingly been given to accurately report the advances made, as well as placing the results into a greater context of the enfolding field. Meanwhile, in the thesis, the text focuses on explaining the results in an as accessible fashion as possible, and how

these results were arrived at in the first place. Hence, the thesis part is more limited in the sense that its scope is more fundamental, but on the other hand more liberty is given to exploring underlying concepts, so at some points it will be more elaborate. A consequence of this is that the reading experience might be quite different between the thesis and the paper parts. In a sense, it is two views on the same results.

Part I

The Friction of Layered Atomically Thin Materials

Chapter 1

Introduction to Friction

Much to the chargin of the physics community, applications of physical theories are not always restricted to idealized conditions [4]. This is the stage for much of this thesis. The purpose of this chapter is to introduce the topic of tribology – the study of friction and lubrication – on a superficial level for readers not familiar with it. We will briefly be covering some general important friction results, but quickly be moving on to friction at the atomic scale. After that we will review how to model and simulate such systems. But first, we shall be giving a motivation to how a theoretical physics PhD came to be largely about friction. A reader familiar with (nano)friction and contemporary simulation techniques can probably safely skip this chapter.

Why you should care about friction

About half of this thesis concerns itself with the matter of friction. Now, you may ask, *"Why should I care about that?"*. We will be supplying three strong reasons why here:

1. Friction is ubiquitous. Unfortunately friction rarely gets the spotlight it deserves – rather it tends to disappear in the shadow of other seemingly

more important everyday phenomena such as gravity, the motion of bodies or the transfer of energy. However, to anyone who doubts that friction has a lesser impact on our everyday lives I riddle you this: *"Look around you, and point at one single thing that would function the way it does regardless of friction."* There are some few examples if you think about it hard enough, but in general, friction is about as influential on our everyday life as most other overarching physical concepts. This might seem fallacious to some people having studied physics, after all, it seems like virtually every physics problem in any text book assumes: *"that there is no friction and no air resistance"* (the last of which, by the way, is also a kind of friction). Truthfully, what they should have written is *"assume friction behaves just the way you need it to in order for the dynamics of this exercise to come out the way you need them to"*, because that is generally what they mean. However, restating that disclaimer every time is pretty cumbersome, so telling students to simply forget about friction is an easier solution – although it comes at the cost of diminishing the importance of friction.

2. Friction has a tremendous impact on modern society. Not only is friction everywhere, but it also affects everything. There are a lot of situations where we actually rely on friction, e.g. friction is imperative for our ability to walk – moreover, in a world without friction Monty Python's *Society for Putting Things Upon Other Things* would quickly be decommissioned, as putting anything on anything is quite impossible without the presence of friction. However, in the study of tribology, we mostly investigate ways of decreasing friction rather than increasing it. This is referred to as lubrication. The reason for this is that there are immense energy losses tied to friction in various (industrial) applications. In fact, recent estimates attribute 23% of the world's total energy consumption to tribological contacts, out of these 3 percentage points are simply wasted as wear. Globally, if modern day tribology technology was implemented, 450 000 million euro and 1 460 million metric tonnes CO_2 could be saved over the next 8 years [5]. If nothing else, those figures go to show that friction is, or should be, of paramount importance to the engineering sciences.

3. Friction is an intriguing problem in its own right. Disregarding the fact that friction concerns everything, and that it has a decisive impact on global economics and energy consumption – which are two pretty hefty omissions – even so, there is still good reason to study friction for the sole sake of its sheer complexity. Friction is a multiscale multidisciplinary phenomenon. So much so that it engages physicists, engineers, chemists, mathematicians, and even the odd biologist. A complete model for friction would have to span approximately nine orders of magnitude – all the way from atomic interactions up to highly macroscopic roughness patterns. Moreover, the dedicated study of tribology as a stand-alone subject is still in its nativity verily, the term tribology was coined only in the 1960s [6], which indicates that there is still a lot to do even on a superficial level.

In summary, you should care about friction because: 1. For better and worse it occurs everywhere, 2. It has a great impact on modern society, and 3. It is a deep and interesting field in its own right. Obviously, we will not cover the field of tribology in its entirety in this thesis, rather we will be focusing on the friction of 2D materials in particular. Friction is a vital part in harnessing the great potential shown by these materials, so whether you are interested in nanomaterials or friction, you will hopefully find something of value in this part of the thesis.

1.1 Friction – on average

As noted in the statement *"Why you should care about friction"*, it is an interdisciplinary, multiscale, and surprisingly elusive problem. There are however a few pointers available, at the macroscopic scale perhaps most notably Amontons' phenomenological laws of friction which read [7]

I The friction force is proportional to the applied load.

II The friction force is independent of the apparent contact area.

III Kinetic friction is independent of the sliding velocity[1].

The first of these laws seems intuitive enough, and it can easily be reproduced by anyone by simply taking for example a hand and put it on a desk and then try and slide it while pushing down to various degrees on the surface. This is all contained in the famous friction formula

$$F = \mu N, \tag{1.1}$$

where μ is a system parameter called the coefficient of friction and N is the normal force.

The other two laws though, might seem counterintuitive at first. Most people would probably agree that pushing a box B_1 with mass m should be less arduous than pushing a box B_2 which also has mass m but is twice the size. While this could be true in some special cases, *in general* it turns out not to be. This comes down to the surface roughnesses of the boxes and the floor, which are stochastic in nature and can be invisible to the naked eye. The friction force is mediated by the actual contact points between two surfaces, and since the roughness is stochastic, the contacts will be localized to where the two surfaces happen to meet. The more contact, the stronger the friction, see figure 1.1. The point here being that the real contact area is not the same as the apparent contact area, and moreover that it is the real contact area that determines the friction, as stated by Amontons' second law. For our hypothetical boxes this means that while we did increase the apparent contact area when going from B_1 to B_2, we at the same time decreased the pressure by the same factor, meaning on average the real contact area will be the same. This pressure dependence also explains our thought experiment with the hand, illustrating Amontons' first law, the harder we push, the more true contact area we get.

In a similar fashion, as we start pushing our imaginary boxes along the floor, the friction will change as the contact areas between the two

[1]This is actually Coulomb's law, but it is often written in conjunction with Amontons' laws.

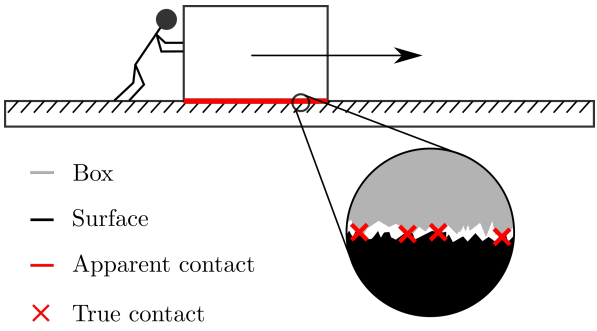
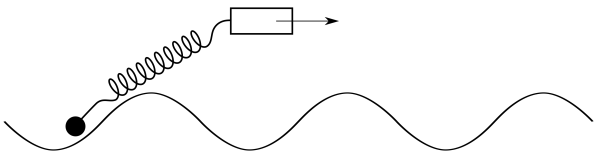


Figure 1.1: An illustration of the true contact area, which is located to isolated asperities.

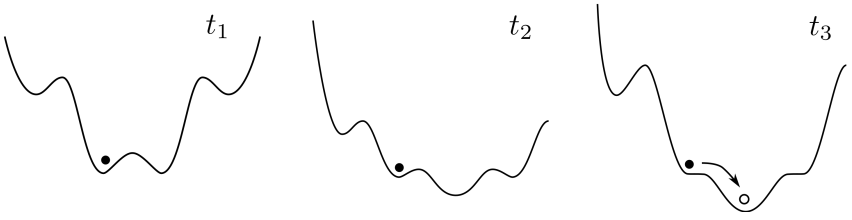
surfaces evolve with time. However, the friction is still dominated by the contact area, which on average will stay the same, meaning that the friction does not depend on velocity, in accordance with the third law. The static friction of objects tends to be higher than their kinetic friction though, since a stationary object has more time to build up a strong contact than a moving one, this is referred to as contact aging, and it is apparent across all realistic length scales [8, 9].

1.2 Nanotribology

This thesis is not particularly concerned with microscopic friction and not at all concerned with macroscopic friction, rather we will mostly keep to the nanoscopic. At this scale Amontons' phenomenological laws collectively break down. The reason for this to a large extent is that the source of friction fundamentally changes. Whereas we macroscopically describe friction as a multi-asperity contact, the contacts are now essentially single asperity, since we are approaching the length scale of individual atoms. As such, we are in need of a new model for single asperity contacts, and the undisputed champion of this is the Prandtl-Tomlinson (PT) model [10, 11, 12]. In this model a hypothetical particle is moving



(a) A schematic diagram of the PT-model.



(b) The potential energy time evolution, with the position of the tip indicated.

Figure 1.2: The PT-model visualized. In (a) we see how a tip represented by a point particle, is being pulled via a spring by a support moving at constant velocity. In (b) the stick-slip process is sketched. First the tip is stuck deep in a potential minimum. Then as the spring is extended, the landscape shifts and the energy barrier shrinks. Finally, the energy barrier disappears, and the tip slips.

on a periodic surface while being pulled by a moving support, see figure 1.2a. The particle will periodically be stuck in potential minima and then slip over energy barriers as the force in the spring overcomes that needed to climb the potential. The resistance to movement (the friction force) can readily be calculated in this model as the time average of the forces on the spring.

Part of the reason for the PT-model's great success is that it is a fairly good representation of an Atomic Force Microscope (AFM)[2], where the

[2]For historical clarity it should be noted that the PT-model has been around a lot longer than the AFM, it is a happy coincidence that the PT-model corresponds so well to AFMs – this was not by design. Also it should be noted that stick-slip motion is *not* inherent to the PT-model, nor to the nanoscale, it occurs across all

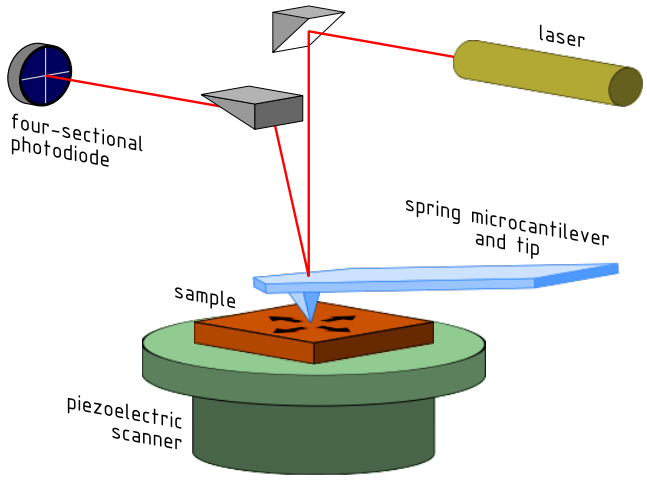


Figure 1.3: A sketch of an atomic force microscope (AFM). The analogue in the PT-model of the tip is the test particle, the substrate is the sample, and the cantilever is the support and spring[3].

periodic surface is some rigid – possibly crystalline – material, the tip is the particle, and the spring and support is the cantilever being pulled by the stage, c.f. figure 1.3. The PT-model is formulated

$$U(x(t),t) = \frac{k}{2}(x(t) - vt)^2 + V_0\left(1 - \cos\left[\frac{2\pi}{a}x(t)\right]\right), \qquad (1.2)$$

where: $U(x(t),t)$ designates potential energy, k is the spring constant, $x(t)$ is the position of the particle, v is the speed of the support, V_0 is the corrugation amplitude (the height of the potential barriers), and a is the lattice parameter (the separation of potential barriers). Furthermore, dissipation is modeled as a viscous damping ($-m\gamma\dot{x}$) term in the resulting equations of motion.

The total potential energy is a corrugated parabola, where the corrugation is suppressed far away from its minimum, this will be important

[3] length scales.

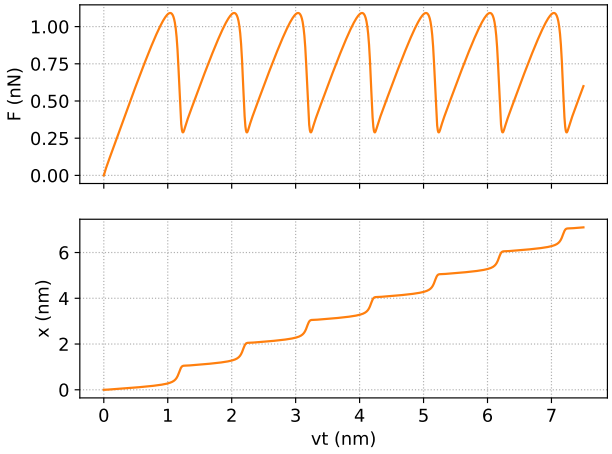


Figure 1.4: A typical force trace as well as tip position as obtained from the PT-model. We see stick-slip motion as the motion of the tip is mostly restricted to small instances of slipping, and otherwise it is sticking.

in the next chapter and is illustrated in figure 1.2b at three different times. A review of the Prandtl-Tomlinson model and its importance to nanotribology can be found here [13].

The resulting lateral force can be calculated as

$$F_{\text{lat}} = k(x(t) - vt) \tag{1.3}$$

and the friction force is the time average of this. A typical force trace is given in figure 1.4.

As we approach the nanoscale, surface corrugation is no longer the only dominant factor determining the friction. At this scale we also have to account for thermal fluctuations [14, 15]. A popular way of handling these perturbations in modeling is to treat them as random kicks to the particle. We shall delve into the details of that in chapter 3, but for now

it is sufficient to note that such kicks will shorten the sticking periods by adding random kicks helping the particle to overcome the potential barrier. This is referred to as thermally activated slips and leads to the phenomenon of thermolubricity, which, as the name indicates, captures how these thermally activated slips constitutes a form of lubricity at the atomic scale. Thermolubricity has been investigated at great length over the last two decades [16, 14], and it has been established that it results in a friction scaling like

$$F \propto |\ln v|^{2/3}, \tag{1.4}$$

where the coefficient of proportionality depends on the temperature. The derivation of this expression as well as a deeper discussion on thermolubricity containing among other things the actual distribution of thermally activated slips is provided in reference [17].

1.3 Simulating friction

Just like in almost every other area of physics, progress in tribology has become increasingly reliant on simulation results over the last couple of decades [18, 19]. There are two main approaches to simulate almost any many-particle system:

1. Simulating some governing equations, e.g. the equations of motion, and calculate the time evolution of the system.

2. Making tiny random changes to the system and using some acceptance-rejection criteria to determine whether to keep the change.

The second of these items is commonly referred to as a Monte-Carlo (MC) type algorithm, even though in principle only a subset of such

[3]Image credit: Grzegorz Wielgoszewski at Wikimedia commons: `https://commons.wikimedia.org/wiki/User:GregorioW`.

algorithms are actually MC algorithms[4]. Regardless, we are not relying on MC algorithms in this thesis and we will not elaborate on them further, the reader who is interested in MC (or MD) algorithms is referred to [20] for a thorough introductory level treatment.

As for simulating the equations of motion, in the case of molecular scale systems – which is what we are investigating in this thesis – this is called a molecular dynamics (MD) simulation. These are often massive scale, hyper-realistic simulations aimed at making an as detailed representation of some physical system as possible. These simulations can have resolution down to individual atoms, and various force fields between every atom are considered. Due to the complexity of these simulations, dedicated MD simulation softwares, such as LAMMPS [21] or GROMACS [22], are often used to implement them. While these simulations can be very detailed and thus can yield a good representation of a physical system, they do also require a vast amount of time and computational resources. The opposite approach to this kind of simulation would then be *simple models*. Here we rely on capturing the relevant components of a system in some clever way, and to make predictions based on this. Simple models are a qualitative way of gaining understanding of a system and how it behaves, and it is computationally cheap as well.

Take AFMs for example. Imagine that we slide an AFM-tip made of silicon over a crystalline NaCl surface and obtain some force trace exhibiting stick-slip dynamics. We could quantitatively recreate this experiment in the computer by using an atomistic simulation. While costly, this should, if all goes well, yield new "experimental" data, to some approximation being equivalent to that of the original experiment. However, we can also represent the system using the PT-model, and calculate the time evolution from the equations of motion at several orders of magnitude of the computational cost. Moreover, a detailed atomistic

[4]This term has been significantly watered down since the formulation of the actual MC algorithm. Nowadays, MC is widely used to describe *any* algorithm significantly relying on some stochastic component, regardless of there even being an acceptance criterion or not.

simulation does only describe the *exact* system it was designed to capture. A simple model can be employed to describe the dynamics of the NaCl system, and any roughly similar system. Furthermore, it is very easy to make parameter studies using a simple model. Also it can be related to other analytical results and tools to further the understanding of the system.

Be advised that there are errors tied to both these methods. These errors come in two kinds: systematic and numerical. The first kind relates to information lost when we set up our system, e.g. in an atomistic simulation we generally do not simulate individual electrons, or we may be disregarding quantum effects[5]. Meanwhile, in a simple model, we can e.g. over-simplify the system and miss to include some important mechanism. Regardless of how we set up our system, the differential equations governing the time evolution will most likely be too complex to be solved analytically, and any attempted numerical solution will be subject to numerical errors. Typically we would look at a system of Euler-Lagrange equations

$$\begin{cases} m_1\ddot{x}_1 = -\dfrac{dU(x_1, x_2, ..., x_n)}{x_1} - m_1\eta_1\dot{x}_1 \\ m_2\ddot{x}_2 = -\dfrac{dU(x_1, x_2, ..., x_n)}{x_2} - m_2\eta_2\dot{x}_2 \\ \quad ... \\ m_n\ddot{x}_n = -\dfrac{dU(x_1, x_2, ..., x_n)}{x_n} - m_n\eta_n\dot{x}_n, \end{cases}$$

where x_i is some dynamic variable, m_i is inertia and η_i is damping, and i is just an index. These equations of motion can then be solved using some numerical time integration algorithm, such as Runge-Kutta or Velocity Verlet, to simulate the time evolution of the system.

[5]One infamous approximation is that the sliding speeds of AFMs in computer simulations (regardless of type) are usually several orders of magnitude larger than in reality. We will return to this later.

Since we are on the nanoscale, regardless of your choice of time integration algorithm, we have to be aware of thermal fluctuations. As noted earlier, this is handled within a framework of random kicks. However, in order for the system not to grow unstable, the kicks need to be counteracted by some damping. This is all captured within Langevin dynamics where

$$-m\ddot{x}_i(t) = -\frac{dU(x_1, x_2, ..., x_n)}{x_i} - m\eta\dot{x}_i(t) + A\xi(t), \tag{1.5}$$

with η being damping and $\xi(t)$ representing random decorrelated kicks scaled by $A = \sqrt{2m\eta k_B T}$ in accordance with the fluctuation-dissipation relation. ξ is a gaussian with $\mu = 0$ and $\sigma = 1$, and furthermore $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$, where δ is a delta function.

Chapter 2

Friction In Layered Materials

In this chapter we will be presenting our model and subsequent results relating to the modified PT-model for layered materials. We will be covering basic terminology about layered materials and expand the PT-model to account for new dynamics introduced in these materials, and then see how this relates to present day research in the field. This chapter summarizes Paper I [1].

2.1 Expanding the PT-model

As successful as the PT-model has been in modeling atomic friction over the past century, there are situations that it does not capture, and that are poorly understood from a theoretical point of view. One such example being to capture the friction in atomically thin layered materials. These systems are similar to the classical PT-systems as discussed in the previous chapter (fig. 1.2a, eq. 1.2), but in this case a sheet consisting of one or more layers of for example graphene is put between the tip and the substrate (see figure 2.1).

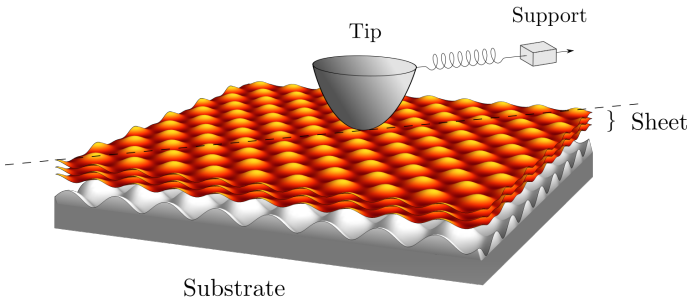
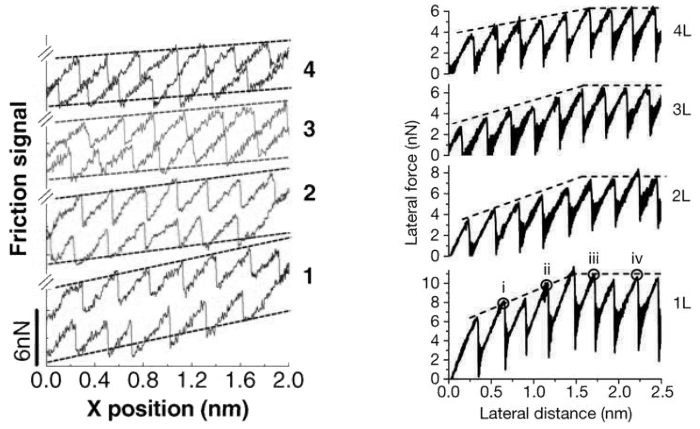


Figure 2.1: The experimental setup studied in this project. An AFM tip is sliding over a sheet made up of some layers of an atomically thin material, deposited onto a substrate. Image from [1].

While both the PT-model and friction experiments on layered systems exhibit stick-slip motion, the experiments also typically exhibit a characteristic initial period of increasing friction (called friction strengthening), whereas the PT-model does not [23, 24]. An example of such strengthening is shown in figure 2.2. This is where our story begins.

Since the first observation of friction strengthening on atomically thin sheets, various authors have speculated as to the origin of this dynamic. Popular suggestions have been that it is related to some kind of bending or puckering of the sheet [23, 27, 26, 28, 29]. Typically, the argument here would be that, as the tip starts to move, it mechanically introduces some out-of-plane deformations to the sheet. The system requires additional energy to increase and maintain these deformations which build up over time until, due to balance of forces, some maximum deformation is reached, effectively cutting off the friction strengthening. However, recently more sophisticated mechanisms involving an evolving quality of the contact has also been proposed. Notably, in [25] Li et al. theorizes and attributes that the local pinning of individual atoms on the AFM tip is what drives the friction increase, arguing that as time evolves the atoms will be deeper pinned, resulting in an increasing resistance to the



(a) Friction strengthening in an experiment [23].

(b) Friction strengthening in a simulation [25].

Figure 2.2: Examples of friction strengthening in experiments and detailed MD simulation. We see multiple characteristic features of friction on 2D-materials here, such as strengthening, layer dependence and a sharp cutoff (although not visible in (a), experiments also report this cutoff, see e.g. [26]).

movement of the AFM tip. The plethora of different explanations available for describing this dynamic have led to some discourse in the field [30]. Regardless of what mechanics give rise to the friction strengthening, ideally, the strengthening would be captured by some simple extension of the PT-model. Let us assume that the strengthening mechanism can be captured by some extra degree of freedom, q. Hereinafter we will refer to q as the sheet distortion, and it is taken to have the unit of length.

We would then have to introduce some energy parameters to capture how the corrugation changes as the sheet deforms as well as internal energy penalties to deforming the sheet, we shall call these κ and ν respectively. Finally, the sheet distortion and the position of the tip should be coupled, since changing one inevitably changes the other, this introduces a phase shift to the periodic term of the PT-model. A candidate potential energy

is

$$U(x(t),q(t),t) = \frac{k}{2}(x(t)-vt)^2 + (V_0+\kappa q(t)^2)\times \\ \times\left(1-\cos\left[\frac{2\pi}{a}x(t)-q(t)\right]\right)+\nu_2 q(t)^2+\nu_4 q(t)^4, \tag{2.1}$$

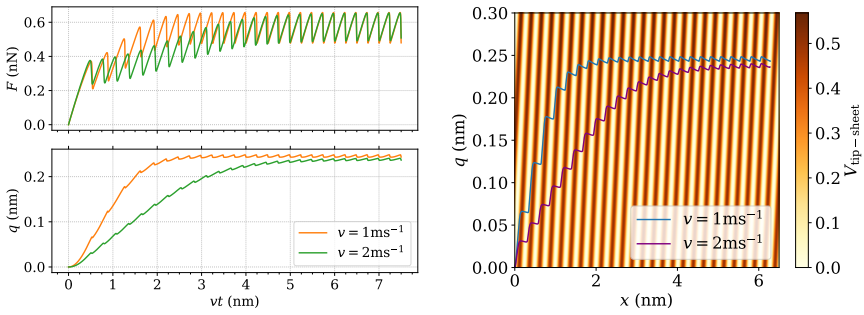
We will return to the nature of these parameters later and why ν is of the peculiar order 4. For now we simply note that it should be an even power because positive and negative distortion should be treated on equal footing, as well as that it has to dominate the corrugation potential to efficiently limit the overall sheet distortion. Limiting the sheet distortion agrees well with the intuitive picture that it corresponds to some out of plane deformation. It would build up gradually as the AFM tip starts to slide over the surface, and eventually saturate at some maximum deformation where it would not be energetically favorable to deform the layer anymore due to a balance of forces between the adhesions and corrugations of the tip, sheet, and surface. This description also agrees well with the idea of an evolving quality of contact among other suggested friction strengthening origins.

Using this potential we obtain the following equations of motion, as described in last chapter

$$\begin{cases} m_x\ddot{x} = -\dfrac{dU(x,q)}{x} - m_x\eta_x\dot{x} & (2.2) \\ m_q\ddot{q} = -\dfrac{dU(x,q)}{q} - m_q\eta_q\dot{q}. & (2.3) \end{cases}$$

This system can then be solved[1] to retrieve the time evolution of the system.

[1]Unless otherwise stated all differential equations in the thesis is solved by Runge-Kutta 4 (RK4) implemented in C++.



(a) Friction force trace (upper), and layer distortion (lower).

(b) The corresponding trajectories in a simplified potential landscape.

Figure 2.3: The dynamics from a first extension to the PR-model. Images from [1].

It turns out that while this simple extension *does* exhibit friction strengthening, it does not quite make it all the way. In particular, it does not exhibit the sharp cutoff seen in experiments and atomistic simulation, c.f. fig. 2.2, nor is it quasi-static (as experiments and simulations have been shown to be) since the friction depends on the sliding velocity. The potential landscape as well as the resulting friction plot of this first extension to the PT-model is given in figure 2.3.

We see a clear asymptotic behavior of the friction here where it slowly approaches its maximum value, there is also a clear dependence on the velocity. The reason the faster tip takes longer to reach its maximum distortion is that the distortion relaxes a little bit during every slip, and traveling at a higher speed, there will be more slips per unit time as compared to the lower velocity case. This would be a reasonable point to interject and ask why we bothered investigating this case at all, if it did not result in anything physical. There are two reasons for this: firstly, it is, in fact, physical, as we shall see in a bit, secondly, we set out seeking a simple PT-like model exhibiting stick-slip motion and friction strengthening, and we have obtained just that. We can now use this as a baseline for developing more accurate model candidates.

Let us ponder the experimental setup for a moment. Comparing figures 1.2a and 2.1 it seems as though the modifications we have made thus far is to allow for the possibility of the periodic surface in the PT-model – or the sheet in the modified PT-model – to deform. While this is all well and good, we have inadvertently disregarded the substrate in our new set up—realistically there would be a substrate under the sheet. There are probably multiple ways of introducing the substrate to this model, but we will use the following[2]

$$U(x,q,t) = \underbrace{\frac{k}{2}\left(x-vt\right)^2}_{\text{Tip-support}} + \underbrace{\left(V_1+\kappa_1 q^2\right)\left(1-\cos\left[\frac{2\pi}{a}(x-q)\right]\right)}_{\text{Tip-layer}} + \\ + \underbrace{\left(V_2+\kappa_2 q^2\right)\left(1-\cos\left[\frac{2\pi\gamma}{b}x\right]\right)}_{\text{Tip-substrate}} + \underbrace{\nu_2 q^2+\nu_4 q^4}_{\text{Layer distortion}}, \tag{2.4}$$

where x is the position of the tip, q is the layer distortion (we have omitted writing out the time dependencies), and the model parameters are explained below. The terms have been labeled to specify which part of the system they capture, see fig. 2.1. Using this expression the friction force can be calculated as outlined in last chapter, c.f. eq. 1.3.

Tip-support term

This term is unchanged from the PT-model. It concerns the tip and captures how the support interacts with the tip. Here k is the spring constant, v is the velocity of the support.

[2]Certainly, the parameters has to be chosen to resemble a real physical system for the outcome of this model to be intelligible. In the spirit of reproducibility, we have settled for parameter values close to those used in reference [25] to describe a mono-layer graphene sheet. The actual parameter values are given in Paper I. We want to stress however, that the model is not sensitive to changes in the parameters, and that it is quite general in this regard.

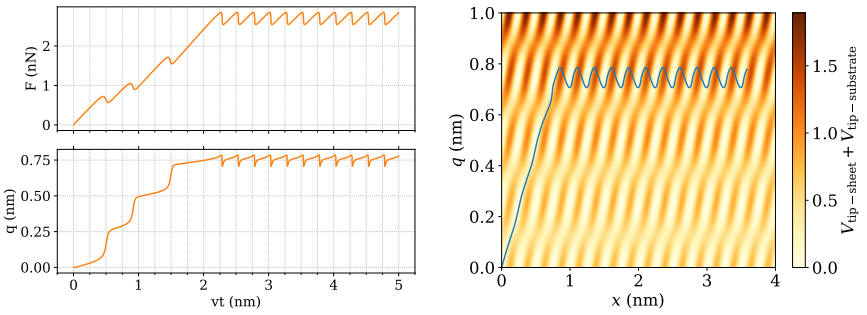
Tip-sheet term
As the tip slides over the sheet, it must overcome the corrugation potential, and the corrugation changes with the layer distortion, this is captured by V_1 and $\kappa_1 q^2$ respectively. The phase shift makes sure that changes in one of the variables directly impacts the other.

Tip-Substrate term
The tip interacts with the substrate through the sheet. This results in similar dynamics as in the Tip-Layer term with (smaller) corrugation and corrugation distortion terms, V_2 and $\kappa_2 q^2$. This term also allows for different lattice parameters for the lattice and substrate.

Sheet distortion term
This term describes how resistant the sheet is to distortion. The energy penalty for deforming the layer must be steep in order to properly limit the distortion, hence we use both the square and quartic terms.

Setting up a system of mono-layer graphene on a commensurate substrate (parameter values found in Paper I) and solving the resulting equations of motion we find the position, layer distortion, friction force trace, and potential landscape[3] as they are given in figure 2.4. These plots are in good qualitative agreement with experimental and simulation results [25, 23]. The friction cut-off is sharp as it should be, and the system is quasi-static and therefore not velocity dependent (not shown in plot), we will see why this is in a moment. From these plots the connection between the layer distortion and friction is clear – so long as the distortion is free to increase, the friction will also increase. Notice how we have not constrained the origins of the distortion further than giving it the unit of length. As far as our model is concerned, the origin of the distortion could in principle be any of the suggestions mentioned earlier (layer deformation and puckering, or evolving quality of the contact), a

[3] Whenever we are referring to "the potential landscape" in this chapter we will be taking about the potential energy minus terms the tip-support and sheet distortion terms. This does not change the qualitative layout of the landscape, but it restricts which minima that are accessible at which times as we shall see in the next chapter.



(a) Friction and sheet distortion. (b) Trajectory in potential landscape.

Figure 2.4: A typical example of how the time evolution of the dynamics look in the modified PT-model. The potential landscape here is just the tip-layer and tip-substrate terms of the potential (eq. 2.4), to increase the contrast. Images from [1].

combination of them, or something completely different.

Finally we note that the substrate can be removed from our model above by setting $V_2 = \kappa_2 = 0$, which eliminates the tip-substrate term. We also note that the resulting potential is almost identical to our first naive extension model (eq. 2.1), meaning that our naive extension was not so naive after all, it is actually a model for a suspended[4] sheet. Figure 2.3 then *could* correspond to the predicted friction trace for a suspended sheet. We write "could" here because there are no experiments to use as reference for strengthening in suspended layered materials. Starting from the reference values for all parameters, in order to get the strengthening for the suspended case, we have to remove the substrate *and* slightly tweak the parameters. If we simply use our reference parameters and just remove the substrate, then we find no strengthening but rather a PT looking force trace. The details of this are presented and discussed in depth in Paper 1.

[4] As in suspended over a cavity – not as in suspended in a liquid.

2.2 The layer-substrate commensurability

Before we conclude that this model successfully captures the force traces correctly in layered systems, we have some introspection to do on what assumptions we have made. Certainly, we have neglected many practical aspects, we will focus on a few that might impact the qualitative behaviour of the model. We already remarked that the model is robust to changes in the parameters, so our specific choice of parameters should be unproblematic. So far we have only investigated mono-layer sheets which are commensurate with the substrate, and moreover we have ignored all thermal effects. We will now focus on unraveling the impact of the first two of these assumptions, because the third one is involved enough that it mandates a chapter in itself (and it is indeed the subject of the next chapter!).

We shall begin by investigating the effect of (in)commensurability between the sheet and the substrate. Let us define what we mean by commensurate, because it will be an important concept going forward. Let a and b be two lattice parameters. Then, two corrugated surfaces with these corresponding lattice parameters are said to be commensurate if $\gamma = b/a \in \mathbb{Q}$. This definition is not very practical however, because in practice, it is impossible to construct a truly incommensurate lattice ratio, since this would correspond to an irrational number. Accordingly, we can rank commensurabilities by finding the smallest integer n such that $\gamma \times n \in \mathbb{Z}$. On a bit of a curious note, it is also possible to rank the incommensurabilities in terms of considering how irrational they are. One can show, using continued fractions, that the most irrational number (in the sense that it is the number worst approximated by a rational number) is the golden ratio $\phi = (\sqrt{5}+1)/2$ [31], this then we shall refer to as the maximally incommensurate case. Notwithstanding, going forward we shall not be referring to explicit commensurabilities, but rather we shall be talking about the related *ratio of the lattice parameters*, γ.

Regard figure 2.5. Here we have plotted the friction traces for four different ratios of the lattice parameters $\gamma = \{1.0, 1.5, 1.6, \phi\}$. Notice how the friction yields four different steady state regimes: 1. Single slip

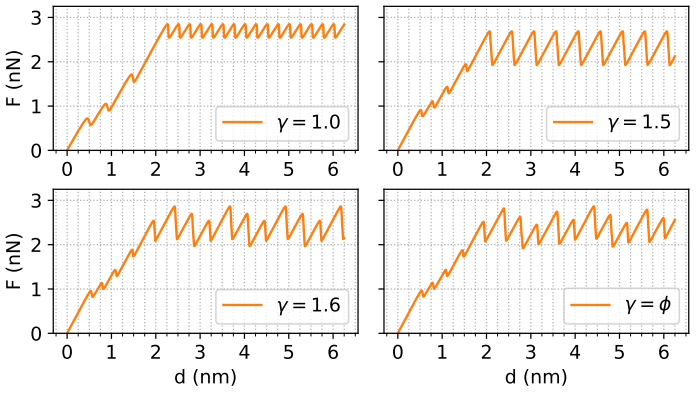


Figure 2.5: The force traces for four mono-layer graphene systems with steet-substrate lattice period ratios: $\gamma = \{1.0, 1.5, 1.6, \phi\}$ respectively, where $\phi = (\sqrt{5}+1)/2$ is the golden ratio.

constant friction, 2. Double slip constant friction, 3. Periodic friction, 4. Aperiodic friction. We notice again that the principal periodicity of those friction traces is given by the smallest integer n such that $n \times \gamma$ is an integer. This can be understood from the Tip-Substrate part of the potential (eq. 2.4), setting e.g. $\gamma = 1.5$ the two corrugations will be in phase every 2 periods, this is then the period of the friction variation. This explains why ϕ, being an irrational number, is aperiodic, as it would have an infinitely long period. We will elaborate on the origin of these friction regimes in the next section, and then among other things understand why $\gamma = 1.6$ has a period of 3 rather than 5. For now though, we note that the resulting friction trace is sensitive to changes in γ.

Widening our perspective, we might want to investigate what happens if we use an irregular substrate. We can emulate an irregular substrate by swapping the periodic substrate in the potential with a set of gaussians separated by distances drawn from a normal distribution. This results in similar overall dynamics with aperiodic stick-slip motion as seen before.

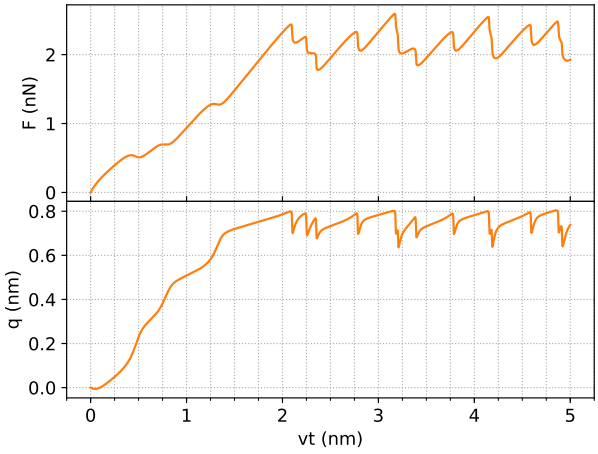


Figure 2.6: When using an irregular substrate, we see that the model still exhibits the same shape and dynamics as before.

The irregular substrate force trace is given in figure 2.6.

We now turn to the mono-layer assumption. While predictions for multilayer systems will be discussed at some length further on in the thesis, for now it suffices to conclude that having multiple layers will increase the resistance to distortion in the system, meaning an increase in the ν_4 parameter. This does not impact the system in any dramatic way.

2.3 The role of the potential landscape topology

Thus far we have shown that the model presented herein adds two crucial components missing from the PT-model as far as layered materials are concerned: 1. Layer distortion, which contributed strengthening, and 2. A substrate underneath the layer, which corrected the physics. At a closer look, it becomes evident that both of those are consequences of the topology of the underlying potential landscape (figs. 2.3b and 2.4b),

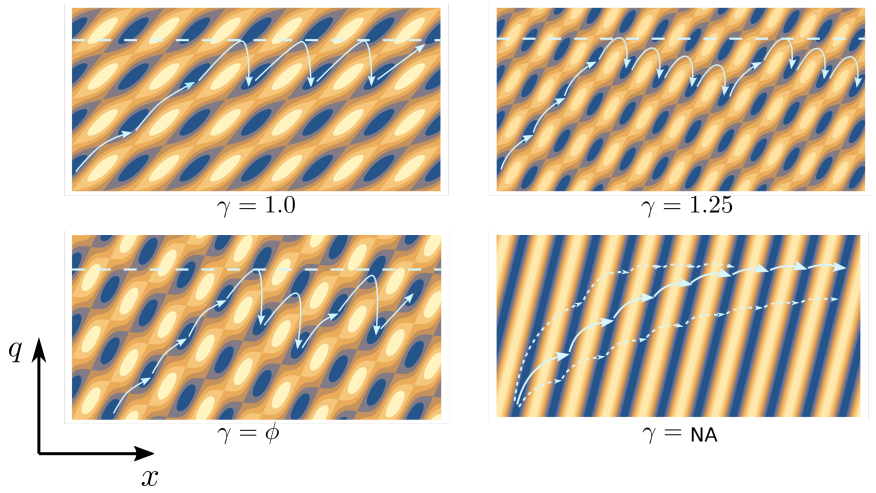


Figure 2.7: Sketches of the potential landscape for various degrees of sheet-substrate lattice ratios. Also a plot of a substrateless case is included. We can see how the friction traces correspond well to the trajectories as they are determined by the potential landscape.

and while this could in principle be captured in one picture, let us make it four for pedagogical reasons as we are not constrained by space, see figure 2.7.

In those pictures we have sketched the potential landscape. That is to say, the parameter values do not correspond to any specific physical system, rather they have been chosen to ensure good contrast in the images to make the features of the landscape visible. Nonetheless, this *is* how the potential landscape looks on a conceptual level in the physical cases too, so we may use these sketches to build an intuition about how the system looks and behaves. A corresponding trajectory has been added in each sketch to illustrate how the tip traverses the landscape.

Looking at the anatomy of the potential landscape, we see how it is made up of interlacing minima and maxima arranged in bands (except in the suspended case). Between those bands are saddle regions with saddle points occurring with the same periodicity as the extrema. We shall

see in a moment how *the path a tip follows in this potential landscape is predetermined by the topology of the potential landscape itself* e.i. it does not depend on any external or dynamic factors. Finally, we may tilt the potential landscape (with respect to the bands) by changing γ.

Already from the fully commensurate sketch we can see why the friction in these systems will be sharply cut-off rather than level off, as well as why they are quasi-static. Assume that the tip starts in some potential minimum, as shown in the sketch. Then, as the support slides along, it will pull the tip in the x-direction. However, as the tip is being pulled, it will distort the sheet, which will push the tip in the q-direction. We have already argued from physical grounds – and we will later analytically prove – that there is a maximum attainable q for any given system in this model. Then, as the tip approaches a saddle region, and if the system is sufficiently far away from $q_{\max}$, the tip will jump over the saddle region into the next band. Since the tip ended up in a higher band, the friction will now be stronger. This is the origin of friction strengthening.

For every band slipped this way, the resistance against continued distortion of the layer will increase, this process will continue until such time that it is no longer energetically favorable to slip into another band due to the increasing layer distortion. At this point, the tip will however still be pulled in the x-direction by the support. The result of this is that the tip will slip into the next minimum in the same band, rather than in the next band. During the slip, the sheet relaxes somewhat, but as the tip sticks in the next minimum, the layer distortion will again increase until the slipping point is reached once more, and thus steady state is reached as the process repeats ad infinitum.

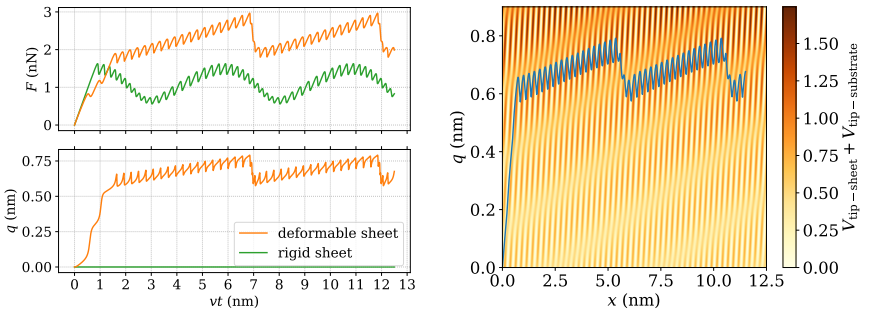
This is why the strengthening is sharply cut-off. In any given band, the friction increases linearly, and the tip will either be pulled up to the next band, or stay in its present band repeating the same motion over, and over again – there is no in-between. Furthermore, this is also what makes this process quasi-static. The trajectory of the tip is given *a priori* by the potential landscape topology. Changing the speed will merely change how quickly the tip slips from one minimum to another,

not how it slips, or to which minimum, within reason[5]. This analysis underscores the importance of being aware of both the sheet and the substrate properties in future simulations and experiments.

Next, let us change γ to investigate the origin of the friction regimes identified at the end of the last section. We already remarked that γ introduces a tilt to the landscape. The effect of this tilt is that, as the tip arrives at its terminal band after the strengthening phase, the next minima might not be readily available to the tip by slipping over the next forward barrier, as a result, the tip will instead follow the tilted band decreasing the sheet distortion significantly. Exactly what happens next depends strongly on the exact alignment of the potential landscape. In the case of $\gamma = 1.5$, as the tip enters the saddle region, it balances perfectly and can slip two lattice periods at once. It then rapidly increases the layer distortion again and ends up back in its initial slipping point, and steady state is reached. In the $\gamma = \phi$ case, the alignment is particularly pathological, and the tip will never reach the same point where it first slipped, this is why the resulting friction trace is aperiodic. In-between these cases we find cases such as $\gamma = 1.6$. In these cases, the tip again follows the negatively slanted band, but eventually, after a series sticks and slips over adjacent bands, the tip ends up in its initial configuration again with $q = q_0$ and $x = n \times x_0$ for some integer n where (q_0, x_0) is the initial slip point, at which point the pattern repeats itself, resulting in periodic friction.

Finally, the last of the sketches in figure 2.7 goes back to the suspended sheet case, where we started out our discussions about layered materials. Removing the layer, and thus removing the second periodic term, the potential landscape shifts dramatically, and there are no interlacing minima, maxima, and saddles anymore. In this case there is no longer any prescribed trajectory for the tip to follow in the potential energy landscape, and we end up with trajectories that are velocity dependent, as seen earlier, meaning that the quasi-staticity is gone. Moreover, since sticking is no longer restricted to specific bands in the potential land-

[5]Certainly, if we increase the speed sufficiently this will not hold, but for reasonable speed changes it will hold.



(a) Friction and distortion, green corresponds to zero distortion.

(b) Trajectory in potential landscape, notice the resemblance to the sketches in 2.7.

Figure 2.8: An example of how moiré patterns show up in our model. Notice that the lateral force period corresponds to friction strengthening in this picture as opposed to figures 2.7 and 2.5. This is because the misalignment is -0.05 ($\gamma = 0.95$), making the rows slanted in the positive direction. A misalignment of +0.05 would correspond to the reverse image, with the lateral force periods corresponding to friction weakening with negatively slanted rows as in the sketches. Images from [1].

scape, the stick slip motion will simply continue until balance of forces is reached and the maximum layer distortion is achieved. This results in an asymptotic approach to the maximum force rather than a sharp cutoff as seen in figure 2.1.

It should be noted that the part about changing γ is well in line with present day research into moiré patterns [32, 33] in atomically thin layers on various substrates. A moiré pattern emerges when there is a slight misalignment between two corrugated 2-dimensional surfaces. The result of such a configuration is that the contact between the surfaces will evolve according to some pattern – called a moiré pattern – because the mismatch of the contact changes in response to the local relative corrugation. An example of how friction, potential energy landscape and layer distortion is represented in our model is displayed in figure 2.8.

Here the green line corresponds to a case where the distortion is fixed at zero. This is included for reference because this is the archetypal moiré pattern found in experiments. Our prediction of how the friction evolves in such a system, in the presence of layer distortion, is supported by experiments as shown in reference [33].

2.4 Analytical estimates

We have now, at some length, investigated how the friction evolves in thin layered materials. As things stand however, actually calculating the friction resulting from some model parameters remains quite cumbersome. We need to simulate the system for sufficiently long to have it entering some kind of steady state, and then we need to measure the average lateral force over several periods. Ideally, we would like to find some closed expression to calculate the approximate friction right away, without having to simulate the system. That is the purpose of this section.

While the full model (eq. 2.4) is fairly involved, the reduced model for suspended sheets (eq. 2.1) is substantially easier to work with from an analytical point of view. We might even suggest that this potential would be a pretty good estimate of the full one, as the dominant limiting factor still is the $\nu_4 q^4$-term, so the contribution from the additional tip-substrate term should be relatively small in comparison. Hence, the steady-state friction in the suspended case should be a fairly good approximation of the steady-state friction in the full model.

Recall that the potential looked like

$$U(x,q,t) = \frac{k}{2}(x-vt)^2 + (V_0 + \kappa q^2)\left(1 - \cos\left[\frac{2\pi}{a}x - q\right]\right) + \nu_2 q^2 + \nu_4 q^4.$$

Also, recall that the friction force was given by $F_{\text{lat}} = (x(t) - vt)$ (eq 1.3), but that this expression actually came from the forces on the spring

pulling the tip. What we are looking for then, is the forces on the spring in the inflection point, just as the tip is about to slip during steady state. At this point, there will be a perfect balance of both x- and q-forces. The q-force, which will be maximized at this point, can be calculated by taking the derivative of the potential with respect to q to obtain

$$\begin{aligned} 2\nu_2 q_{\max} + 4\nu_4 q_{\max}^3 + 2q_{\max}\kappa_1 \left[1 - \cos\left(\frac{2\pi}{a}(x - q_{\max})\right)\right] - \\ -\frac{2\pi}{a}(V_0 + \kappa_1 q_{\max}^2)\sin\left(\frac{2\pi}{a}(x - q_{\max})\right) = 0 \ . \end{aligned} \tag{2.5}$$

Furthermore, at the inflection point,

$$\frac{\partial^2 U}{\partial q^2} = 0. \tag{2.6}$$

We note that this equation gives two constraints

$$\cos\left(\frac{2\pi}{a}(x - q)\right) = 0 \tag{2.7}$$

and

$$\sin\left(\frac{2\pi}{a}(x - q)\right) = 1. \tag{2.8}$$

Using these conditions we retrieve the following polynomial equation

$$2\nu_2 q_{\max} + 4\nu_4 q_{\max}^3 + 2\kappa_1 q_{\max} - \frac{2\pi}{a}(V_0 + \kappa_1 q_{\max}^2) = 0 \tag{2.9}$$

for the maximum distortion $q_{\max}$ attainable by the system. This third order polynomial only has one real root, which then is the sought maximum distortion. Going back to eq. 2.6, we calculate the maximum force as

$$F_{\text{lat}}^{\max} = \frac{2\pi}{a}(V_0 + \kappa_1 q_{\max}^2). \tag{2.10}$$

Summarizing, this expression gives the maximum lateral force. We can relate this estimate to the modeled friction by simulating the friction for some set of parameters. In figure 2.9 we compare the estimated friction to that predicted by the model for an incommensurate system while changing ν_4 – the reason we chose this parameter in particular will become apparent in the next section. We note that the correspondence is very good, and conclude that our estimate is sound.

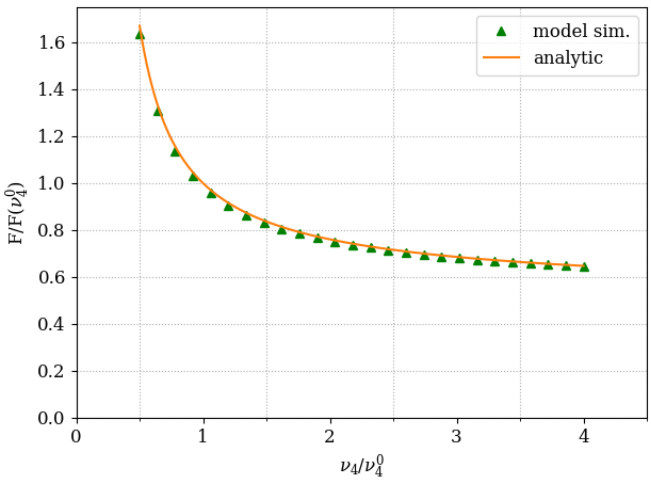


Figure 2.9: Friction as a function of ν_4 as simulated by our model and estimated by our closed expression for the steady-state friction.

For reference we wrap up this section by plugging in the solution to eq. 2.9 into eq. 2.10, and write out the full analytical solution

$$q_{\max} = \frac{2\pi}{a}\left(\kappa_1\left(\frac{p_0^{1/3}}{12\times 2^{1/3}\nu_4} - \frac{p_6 - p_5}{6\times 2^{2/3}\nu_4 p_0^{1/3}} + \frac{\pi\kappa_1}{6a\nu_4}\right)^2 + V_0\right) \tag{2.11}$$

where,

$$p_1 = \frac{144\pi\nu_4\kappa_1\nu_2}{a} \tag{2.12}$$

$$p_2 = \frac{864\pi\nu_4^2 V_1}{a} \tag{2.13}$$

$$p_3 = \frac{144\pi\nu 4\kappa_1^2}{a} \tag{2.14}$$

$$p_4 = \frac{16\pi^3\kappa_1^3}{a^3} \tag{2.15}$$

$$p_5 = \frac{4\pi^2\kappa_1^2}{a^2} \tag{2.16}$$

$$p_6 = 24\nu_4(\kappa_1 + \nu_2) \tag{2.17}$$

$$p_0 = p_2 + p_4 - p_1 - p_3 + \sqrt{(p_2 + p_4 - p_1 - p_3)^2 + 4(p_6 - p_5)^2} \tag{2.18}$$

note that $q_{\max}$ from eq. 2.9 then is the square part of this expression.

2.5 Multiple layers

Significant research interest has been given to multilayered systems over the past decade [24, 23, 25, 34, 35]. One reason for this scrutiny is that this problem is both intriguing from a theoretical point of view and shows great promise for applications. When increasing the number of stacked layers in these systems, the friction does in fact decrease, which might seem counterintuitive at first. According to our model, increasing the number of layers in the sheet, corresponds to increasing

the ν_4 parameter, the physical interpretation being that it is harder to deform a sheet the more layers it consists of. Furthermore, as the sheet distorts, the average distortion in each layer will be about equal, meaning that there is good reason to believe that ν_4 scales linearly with the number of layers.

In fact, we have inconspicuously already shown that the model captures the dynamics of decreasing friction with increasing number of layers. This is evident in fig. 2.9, where friction decreases as ν_4 grows.

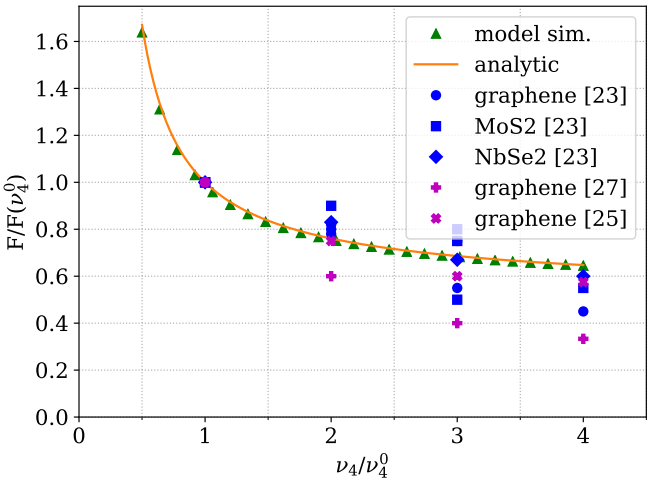


Figure 2.10: Friction as a function of the number of layers in the sheet as predicted by our model (green), the analytical estimate of eq. 2.10 (orange), and experiments (blue) as well as computer simulations(purple). Image from [1].

There is a lot of experimental data available on how the friction changes with the number of layers. As a trial by fire, we may compare the friction as predicted by our model to this data. The result from such a comparison is given in figure 2.10. In order to make the results more comparable, this figure uses normalized friction to compare the relative friction decrease as the number of layers increases. Here we have inferred that $\nu_4(n) = n\nu_4$ where n is the number of layers. Experiments have

consistently shown a decrease in friction of about 20% when going from one layer to two layers, the data going to more layers than that is more sparse and prone to larger errors, but the general trend is as shown in the figure 2.10. In our model, going from 1 to 2 layers corresponds to a friction decrease of 24%, which is well within experimental limits.

2.6 Discussion and outlook

In this project we endeavoured to – with origin in the Prandtl-Tomlinson model – develop a model to capture the peculiarities of atomic friction in layered materials. We devised the model as presented in equation 2.4, and concluded that it captures: friction strengthening, friction cut-off, layer dependence, strengthening in moiré patterns, the steady state friction force, friction with and without a substrate, driving mechanism for all the listed here. All in all, this simple model has been astonishingly successful in describing and predicting a score of experimental and simulation data collected by the community over the last decade.

Perhaps the biggest immediate contribution from this model to the field is that it resolves the long standing conflict regarding seemingly contradictory conclusions drawn about the origins of friction strengthening and layer dependence as proposed by various authors since the first observation of the phenomena. This model provides a framework which fits all those suggested origins – the discourse is moot, because everyone seems to be right. According to our model, layer distortion is a very general concept, whose role can be filled by a multitude of different degrees of freedom at different length scales, possibly even an interplay of several at the same time.

While this calls for some celebration in its own right, there is a lot more to be done in the context of this model. So far we have been very focused on utilizing our model to explain various existing friction data. However, given the simplicity of the model, it should be fairly easy to adopt it to more hypothetical situations to make predictions for new physics to be found in the field. One such example that shows much promise

for potential applications is friction tuning using moiré patterns, see e.g. [36]. We are hopeful that our model can serve to further empower theoretical studies of friction on atomically thin layered materials.

Chapter 3

Thermal characteristics

In this chapter we will address an important aspect of nanoscale friction that was not considered in the previous chapter: The effects of thermal noise on the friction of layered materials. We shall see that the ramifications of introducing thermal fluctuations are relatively small on the overall friction behaviour outlined in the previous chapter, but that it is of substantial importance to how relaxation might occur after stopping the AFM tip in these systems because of the now two dimensional landscape. This chapter covers Paper II.

3.1 Vanilla thermal activities

On a superficial level, the thermal effects on friction in layered materials is similar to that generally observed on the atomic scale [37, 38]. As such, we can directly import much of the considerable knowledge and tools that has been developed in this area with only minor modifications to draw conclusions about what behavior to expect in general. We conceptually introduced Langevin dynamics in Chapter 1, we will reintroduce it here in a more concrete way in terms of the x and q variables. The equations of motion read

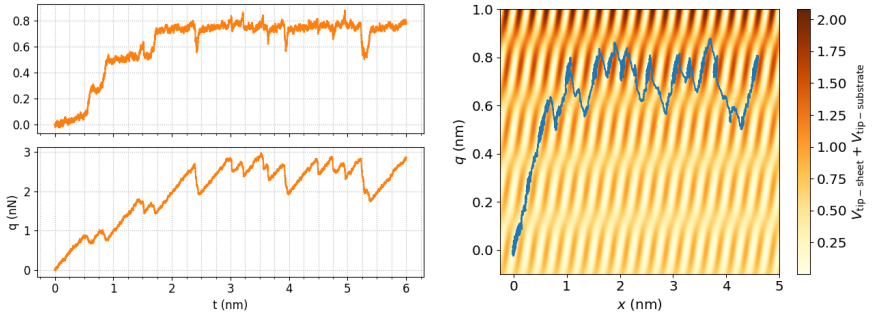
$$-m_x\ddot{x} = -\frac{\partial U(x,q)}{\partial x} - m_x\eta_x\dot{x} + A_x\xi_x(t) \tag{3.1}$$

$$-m_q\ddot{q} = -\frac{\partial U(x,q)}{\partial q} - m_q\eta_q\dot{q} + A_q\xi_q(t)\ , \tag{3.2}$$

where m_x and m_q are the effective masses of the tip and the distortion, η_x and η_q are the viscous damping coefficients, and the prefactor $A_{x,q} = \sqrt{2k_BT\eta_{x,q}}$ together with the gaussian noise ξ_x and ξ_q with unit standard deviation, zero mean, and $\langle\xi(t)\xi(t')\rangle = \delta(t-t')$ captures the thermal noise.

This is the Langevin dynamics used to introduce thermal fluctuations into the system. It is assumed that the fluctuations can be represented as uncorrelated, momentaneous kicks of random magnitude and direction, this is the third term in the equations of motion. The uncorrelated part is taken care of by the last statement above, whereas them being random stems from the gaussian noise. The prefactor to the gaussian noise then naturally becomes a scale factor giving the amplitude of the thermal fluctuations which naturally depends on the temperature. However, simply introducing random kicks to the system is potentially not unproblematic. We have added a channel for increasing the energy of the system, but without also having a corresponding channel to dissipate the energy, the energy will divergently increase with time. Fortunately, by construction, the PT model already contains a viscous damping term to serve this purpose.

Using these equations of motion we can analyze the modified PT model's behaviour in the presence of thermal noise. For comparison, the standard evolution of friction and layer distortion was presented in figure 2.4. The thermal version of the corresponding standard evolution is given in figure 3.1, along with its path in the potential landscape. There are immediate similarities between the two cases: friction strengthening is persistent in the presence of noise; there are clear stick slip signatures; the system still spends the most time in or near minima. There are however also some differences: The average friction is slightly decreased, i.e.



(a) Force trace and q evolution for a typical system realization.

(b) The parametric path as taken in the underlying potential landscape.

Figure 3.1: Using standard parameter values, the system typically evolves like so. After the strengthening regime, the system enters a steady state where the system spends most of its time around the same q-value, and where the friction over time is about constant.

we have thermolubricity as expected; we observe "double slips", where the tip overshoots a barrier and slips two lattice periods; sometimes the tip ends up in a lower row of extrema before returning to the top row; and the signal-to-noise ratio is large enough that the pattern in the q-steady state is mostly gone.

Most of these effects are expected from contemporary understanding of thermolubricity and the amplitude thereof [16]. However, our new degree of freedom, q, does add some new intriguing complexity. In an ordinary PT system, the amplitude of thermolubricity is determined by how the expected slipping point is shifted by thermal noise [17]. A layered system allows for slightly more rich dynamics, since not just *when* the tip slips is important, but also where it slips *to*. The possibility of slipping into a lower region of the potential landscape increases the friction reduction due to thermolubricity and de facto decreases the effect of strengthening as well.

This captures the basic thermal behaviour of the model, and we can conclude that it agrees well with what has been observed in experiments

and simulations [25, 23]. However, there is an elusive crux hiding in plain sight. In the PT-model, when an AFM tip is stopped, the system intuitively relaxes as the tip is pushed towards a global potential minima by random kicks. It is not entirely obvious that this process is as straightforward in a layered system. It seems trivial enough that the tip should be driven into a global potential minimum in the x direction by the same argument it happens in the PT-model, however how q behaves is more subtle, moreover, seeing as though x and q is coupled, we expect a more vivid thermal relaxation pattern. This is what we will spend the remainder of this chapter investigating[1].

3.2 It is all about rates

Thermal relaxation happens due to spontaneous transitions from a higher energy state into a lower one, mediated by thermal fluctuations. The rate pertaining to such an event occurring is called a *decay rate*, and it can be modelled by a corresponding *Kramers rate* [39]. This rate has been widely influential on the theory of reaction rates in chemistry [40], and that is not a bad place to start in order to understand the mechanics of the decay rate. Here, typically one would regard a situation where some reactants have to overcome an energy barrier to form a product with lower internal energy, this is illustrated in figure 3.2. In the left diagram, the separated reactants correspond to a system of higher energy, however, in order to form a complex with lower energy, an energy barrier ΔE must be overcome. Since the system is exposed to a thermal bath – i.e. its surroundings – the reactants will jiggle in their potential minima. If the energy barrier is sufficiently low, at some point, thermal kicks will eventually bring the reactants close enough that they can form a product, that is, thermal kicks will push the system over the barrier. Now the system is in a more energetically favourable state.

[1] In Paper II we further discuss some practical aspects of the thermal behaviour that are of particular interest to experimentalists. In this summary we will however focus on the theoretical development of our model, so we cut this practical part out and refer the reader to the paper for a rigorous treatment.

However, depending on the system in question, there might be – and in this diagram there certainly is – a probability that thermal kicks could push the system back over the boundary, by breaking up the low energy complex. The rate at which this transition is happening is captured by the Kramers rate.

We can describe the oscillations in a potential well with an effective frequency given by

$$\Omega = \sqrt{\frac{\partial^2 U}{\partial x^2} \Big/ m}\,, \tag{3.3}$$

here x is the reaction coordinate of the system, and m is the effective mass. The Kramers rate, and by extension this expression too, are very general results that are applicable to a wide range of systems, the effective reaction coordinate might for example be an external field, in which case the corresponding effective mass, or inertia rather, might be capacitance [41]. Deriving the Kramers rate from this expression is fairly straightforward and involves solving the corresponding Fokker-Planck equation, the derivation is enough and not of substantial material interest in itself, that we are omitting it here. There are however many accounts available that contain the derivation, such as a rather verbose approach here [42]. In terms of this Ω, the Kramers rate is given by

$$\frac{1}{\tau} = \frac{\Omega_A \Omega_B}{2\pi\eta} e^{-\frac{\Delta U}{k_B T}}, \tag{3.4}$$

where η is the damping, ΔE is the energy difference between the two states, τ is the average escape time, and $k_B T$ is the thermal energy.

Typically, the kind of reactions well described by the Kramers rate are reversible thermally driven reactions. One such example would be protolysis in weak acids, e.g. solving carbonic acid in water: $H_2CO_3 + H_2O \rightleftharpoons HCO_3^- + H_{30}^+$. Each of these states correspond to a potential well, as in figure 3.2. Simplified, the depth of each well is determined by the energy of either state, and the barrier height by how much energy is

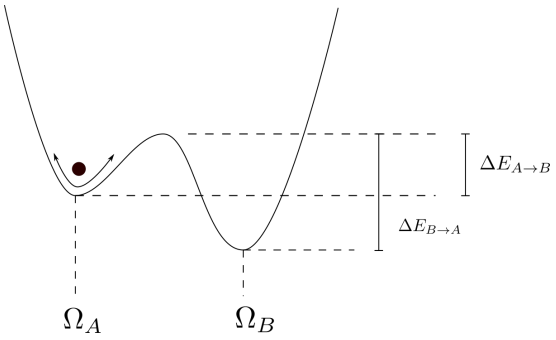


Figure 3.2: A double well potential energy system. Because of thermal fluctuations, the system will be found in different states at different times, the prevalence of a state can be calculated from the Kramers rate and depends on the barrier height, ΔE.

needed for the reaction to occur. The more thermal energy in the system, the more kinetic energy the molecules have, and thus they are more likely to overcome the energy needed for the reaction to happen, which increases the reaction rate, up until some point where other factors we disregard here come into play.

In this thesis, we are however not particularly interested in the chemistry of things. Moreover, as we have previously discussed, the potential landscapes we are working with are significantly more complex than the simple one in figure 3.2, for one thing, they are multi-levelled rather than just having two metastable states.

Worse than that, our potential landscape is two dimensional, which makes things considerably more complex. Let us try and understand some of the implications of this. Assume for a moment that we are looking at some simple regular two dimensional landscape, with a particle originating in some high-energy local minimum, then there are eight principle escape routes available to a particle – four horizontal or vertical, and four diagonal (cf. fig. 3.3). This figure is not a realistic representation of how the potential landscape of our model looks like (we will provide such a figure later) it is merely meant to illustrate qualitative

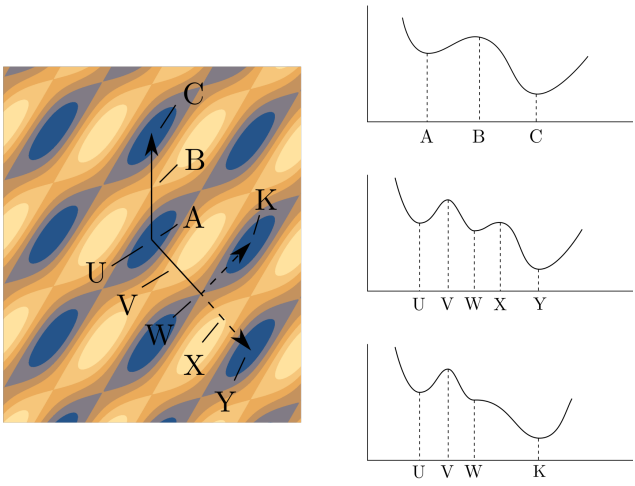


Figure 3.3: Escape paths in a simplified two dimensional potential landscape. The diagrams show the one dimensional sections following the lines in the landscape. We can see how there are multiple escape paths available to a particle starting at A.

aspects of at all having a two dimensional landscape.

In the horizontal and vertical directions the escape path is trivial, and we can in principle use a one dimensional Kramers rate as an approximation to calculate the escape rate. However, in the diagonal directions there is a caveat we need to address. As we diagonally escape from the minima labelled U in the figure, we arrive at the saddle point W (we shall be referring to these as *transition states* later on). From this point, the particle could continue on a path to minimum Y, but it could also follow a channel leading up to the minimum K. Which of these paths will actually occur depends on a range of factors we shall investigate in the context of our model over the course of the coming sections.

This is an indication as to why trying to reduce this two dimensional problem into a set of simpler one dimensional problems is not likely to succeed. Hence, we will need to generalize equation 3.4 to two dimensions. However, the way to properly accomplish that is a paper in itself

[43] and mandates a thorough introduction to transition state theory, so we will simply give the expression here and comment on the extensions from the one dimensional Kramers rate which we have already provided some background to. The higher dimensional Kramers rate for going from one minimum to another via a saddle point reads

$$\frac{1}{\tau} = \frac{\lambda_+}{2\pi}\sqrt{\frac{\det \mathcal{E}_m}{|\det \mathcal{E}_s|}}\,\exp\left(-\frac{\Delta U}{k_B T}\right), \tag{3.5}$$

where $\mathcal{E}_i$ are the Hessian matrices of the minimum and the saddle point respectively and λ_+ is the small displacement exponential growth rate of the saddle point, we explain these below.

Here we have introduced the Hessian matrix, which is just a matrix of the second order derivatives, meaning

$$\mathcal{E} = \begin{pmatrix} \dfrac{\partial^2 U}{\partial x^2} & \dfrac{\partial^2 U}{\partial x \partial q} \\ \dfrac{\partial^2 U sing}{\partial q \partial x} & \dfrac{\partial^2 U}{\partial q^2} \end{pmatrix}, \tag{3.6}$$

evaluated at the maxima and saddle points respectively. The small displacement exponential growth rate is the rate at which a particle leaves the transition state. It is calculated as the largest positive eigenvalue of the dynamic matric $\mathcal{M}$ defined by the equation

$$\begin{pmatrix} \dot{x} \\ \ddot{x} \\ \dot{q} \\ \ddot{q} \end{pmatrix} = \mathcal{M} \begin{pmatrix} x - x_0 \\ \dot{x} \\ q - q_0 \\ \dot{q} \end{pmatrix} \tag{3.7}$$

at a small displacement from the saddle points with coordinates (x_0, q_0). This equation can be solved from the equations of motion (eqs. 3.1, 3.2). Two of the resulting equations are trivially given by $\dot{x} = \dot{x}$ and $\dot{q} = \dot{q}$. Furthermore, we know that

$$\ddot{x} = \frac{F_x}{m_x} - \eta_x \dot{x}, \tag{3.8}$$

We can Taylor expand the force around the point where we need to find it as

$$F_x(x_0, q_0) \approx (x - x_0)\frac{\partial}{\partial x}F_x + (q - q_0)\frac{\partial}{\partial q}F_q, \tag{3.9}$$

and the force is commonly known to be the negative derivative of the potential, so by combining 3.8 and 3.9 we have

$$\ddot{x} = -\frac{1}{m_x}\left(\Delta x \frac{\partial}{\partial x}\frac{\partial}{\partial x}U(x,q) + \Delta q \frac{\partial}{\partial q}\frac{\partial}{\partial q}U(x,q)\right) - \eta_x \dot{x}. \tag{3.10}$$

But these second order derivatives are just components of the Hessian matrix. Repeating the same process for $\ddot{q}$ and gathering everything, we can finally give the expression for the dynamic matrix $\mathcal{M}$ as

$$\mathcal{M} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{m_x}H_{xx} & -\eta_x & \frac{1}{m_x}H_{qx} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{m_q}H_{xq} & 0 & \frac{1}{m_q}H_{qq} & -\eta_q \end{pmatrix}. \tag{3.11}$$

3.3 The thermal relaxation of layered materials

Based on the last section, we are going to simulate how our model relaxes due to thermally activated decays into lower energy states. Since these systems experience friction strengthening, and the steady state friction is reached only after q has reached its limit, we will have to let the system reach this steady state regime before we stop and let it relax.

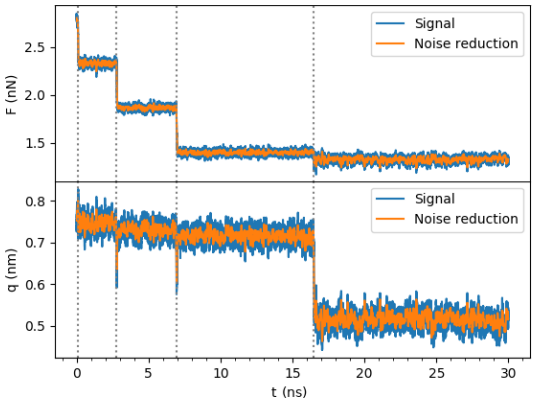


Figure 3.4: An example of how decays are numerically identified and characterized. Dashed lines indicate decay points.

However, due to the stochastic nature of thermal fluctuations, we do not know how long it will take for the tip to reach this point, and even if we did, we do not know in which state it is when we stop the simulation. It could be slipping, sticking, be at a high or low point in the potential landscape, etc. To overcome all these issues we have chosen to prime the system. We did this by numerically obtaining all the dynamic quantities of the system at the first slipping point after strengthening concluded *in a system without thermal noise*. We then save this as the initial state of the system, and in each simulation start from here while artificially "switching on" the thermal noise and subsequently let the system relax. Decays were identified with the algorithm outlined below.

The algorithm for identifying decays is made much simpler by the fact that we know the exact positions of the minima of the potential landscape[2]. Rather than having to develop an algorithm that finds generic decay events, and calculates as well as logs relevant quantities pertaining to these, we can settle for detecting decays in certain predefined

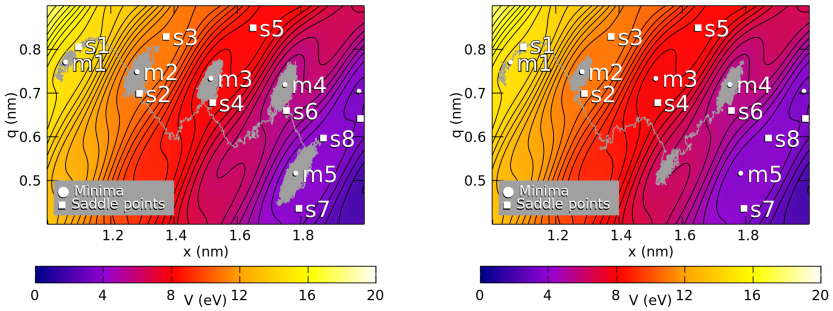
[2]Unfortunately, we only realized this after already having developed an algorithm to detect generic slipping events in stick-signals. That algorithm is omitted here however for the sake of brevity.

regions, because we know that this is where the decays will occur. It is known from construction that the system will start out in the minimum m_1, as indexed by table 3.1 (at the end of this chapter), and also that the system should decay more or less instantaneously. Since the simulations start from a delta distribution rather than a normal distribution we also know that the data initially is correlated across all simulations. On account of this, we eliminate m_1 from the data, since it is problematic. After spending some time in m_1, the tip will decay into some other minima. We detect which one by continuously monitoring three simple criteria:

1. Is the tip closer than some threshold value to any minimum?
2. Is it the same minimum as the last time step?
3. Has the minimum been stable for a sufficient amount of timesteps?

Where the last two conditions ensure that we log actual decays and not just strong thermal fluctuations. If the answer to all of these conditions is "Yes", then we log a new location for the tip and the timestamp for the decay is calculated as the total time elapsed minus the time elapsed since the previous decay. In this way, we always know which minimum the tip decayed from and to, as well as when the decays happened.

Now, let us look at some actual simulation data to get an idea about what we are dealing with here. We have supplied some typical example data in figure 3.4. In this figure the lateral force and sheet distortion are given over time. We have applied a running mean of 1000 points, because it helps with the data analysis, especially since the q signal is quite noisy. The dashed lines indicate decays. We observe how friction decreases in discrete jumps, and that sticking times are increasing between every jump. This is to be expected, since every time an energy barrier is overcome the tip will decay into a state of lower energy. There are however two ways of minimizing the energy, giving some richness to the potential energy landscape: we can move the tip closer to the support, or we can decrease the sheet distortion. The minima are arranged such that there are x decays and q decays labeled by which quantity that changes the most, in this plot we can see how most decays are x decays.



(a) A decay sequence through $m1 \rightarrow m2 \rightarrow m3 \rightarrow m4 \rightarrow m5$.

(b) A decay sequence through $m1 \rightarrow m2 \rightarrow m4$.

Figure 3.5: Two examples of decay sequences superimposed on the potential landscape. The first of these sequences is the modal path, and the corresponding data is also given in figure 3.4. The second case is supplied to illustrate how the potential landscape also admits for more complex decay patterns.

In particular, this decay path corresponds to (again in the labelling convention of table 3.1) $m1 \rightarrow m2 \rightarrow m3 \rightarrow m4 \rightarrow m5$. We have also plotted this data on the potential landscape in figure 3.5a to help relate the data to the topology.

It is instructive to keep in mind that this is only one possible realization out of multiple. The potential landscape is the same across all simulations because we use an identical simulation setup, however, the decay path taken will not be the same, or even deterministic. Another example of a decay path superimposed on the potential landscape is given in figure 3.5b. This time the decay sequence is $m1 \rightarrow m2 \rightarrow m4$. It looks like the tip spends time in a hypothetical minimum below $m3$, however, that is not the case, this region does not contain a stable minimum. What is really happening is that the tip is spending time in the vicinity of a recently vanished minimum in a two dimensional analogy to the one dimensional evolution described in figure 1.2b.

The full potential landscape with all minima and saddle points is given

in figure 3.6, here the dashed lines are the maximum layer deformation as calculated in Chapter 2 and the position at which the support was stopped ($r = 2.448$ nm) respectively. Take a moment to compare this potential landscape to the ones in figures 2.4 and 2.7. A clear difference between these landscapes is the elliptic pattern in this new plot. The reason for this is that neither of the former potential landscapes included the sheet distortion term nor the tip term. These two terms will warp the landscape in the q and x directions respectively. The layer distortion introduces a global gradient symmetric around $q = 0$, the spring similarly introduces a gradient symmetric around the support position vt, these two gradients combined constitutes the elliptic pattern. It is imperative to acknowledge that the previously presented potential landscapes are not rendered erroneos in the face of this new one, rather these are different views of the same landscape. The old landscape highlights the set of principle minima available to the tip, those minima are still present in this plot. However, they are warped by the gradients, much like the simple sinusoidal substrate is warped by the spring in the one dimensional PT-model, c.f. figure 1.2b.

In this figure, the minima and saddle points of the potential landscape have been indicated and labelled. The tip will tend to decay from one minimum to a neighbouring one through a connecting saddle point, the reason for this being that it is easier to transition via the saddle point than going straight for the minimum. A helpful analogy is to imagine the potential landscape as being made up of maxima, minima, saddles and *channels* leading into minima. An example of a channel would be the connection between $s4$ and $m4$, this is perhaps easier visualized in the former reduced landscape if figure 3.3, where we can see a channel leading from the saddle W to the minimum K. The set of minima in figure 3.6 is the full set of minima, there are other would-be minima as given by the underlying potential landscape, however, these are made inaccessible by the present support position and layer distortion.

Based on this potential landscape and using the Kramers theory outlined in the previous section, we may then calculate expected lifetimes and decay paths for the system as it relaxes. The most probable paths are

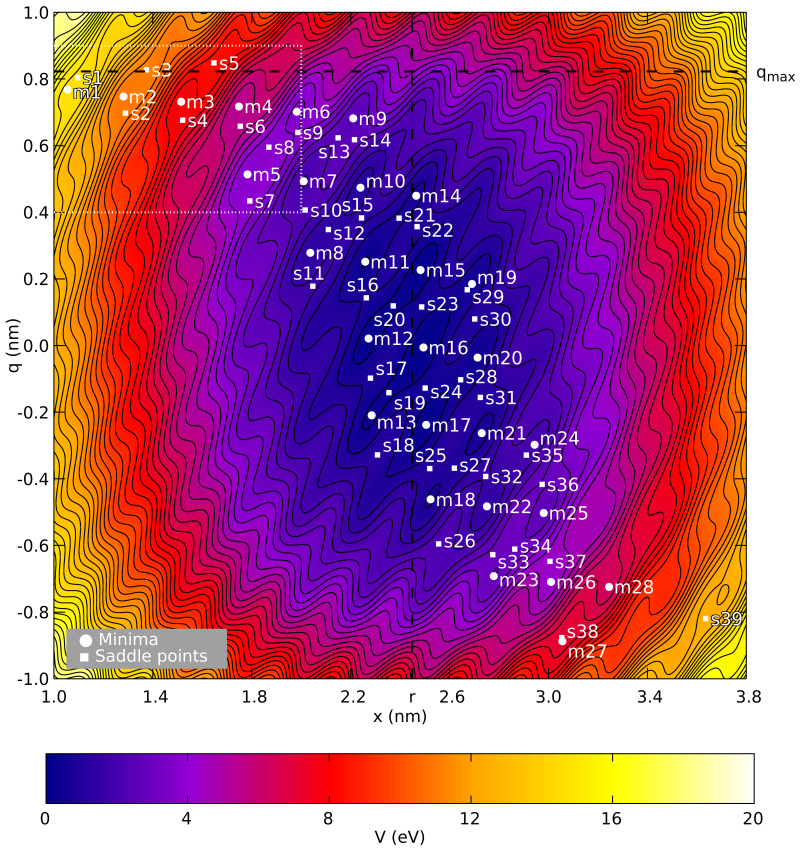


Figure 3.6: The full potential landscape, including all terms in the potential energy at a support position of $r = 2.448$ nm. All available minima have been indicated, as well all the saddle points. The dashed box corresponds to the view box of figure 3.5.

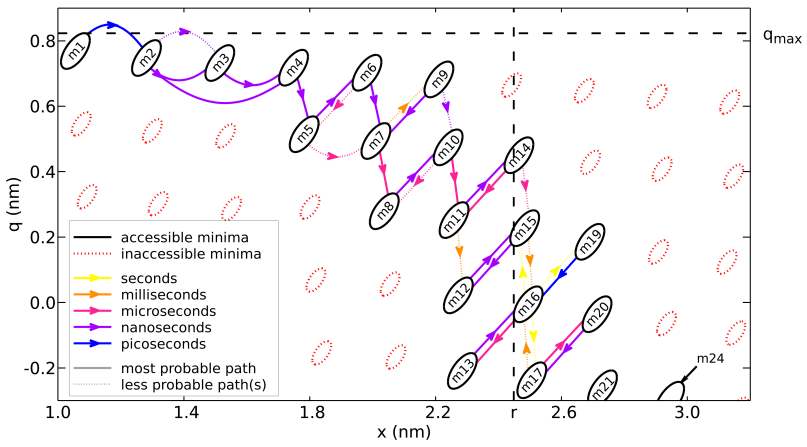
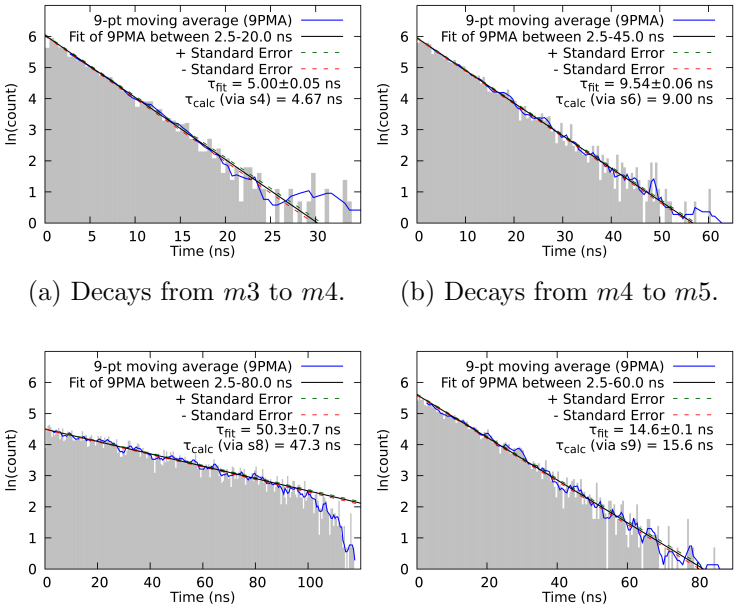


Figure 3.7: A schematic represntation of a selection of decay paths likely to be observable in experiments. The decays has been colour coded according to their time scale.

schematically represented in figure 3.7. From this diagram we can see how there are two phases present. Initially there is a more or less clear decay path with only minor deviations, whereas in more advanced stages of the decay chain we observe reciprocal transitions between pairs of minima. This does not mean that the tip will be trapped in this bilateral state—given sufficient time it will stochastically continue its descent towards the global minimum. The simulations we made in this project were not long enough to reliably verify this back-and-forth behaviour, but the modal path up until minima eight was verified computationally. Do note once again that this is a stochastic system, and this modal path is not the only path that will be taken by the tip, the full width of paths is expected to be much more diverse.

From simulating 10k decays over 120 ns we collected statistics and verified the calculated lifetimes of some selected popular decays. This data



(a) Decays from $m3$ to $m4$.

(b) Decays from $m4$ to $m5$.

(c) Decays from $m5$ to $m6$.

(d) Decays from $m6$ to $m7$.

Figure 3.8: Histogramming the decay times for some selected decays. The fitted decay times can be compared to theoretical values calculated from kramers theory.

is presented in figure 3.8. Here we have binned the logarithm of the observed lifetimes of a particular decay, such as $m2 \rightarrow m3$ in the first figure, the slope of the histogram gives the corresponding decay rate. The fit is calculated with a simple least squares procedure. These decay times can then be compared with those in table 3.1, to get an idea about how accurate the calculated values are. Taking the $m2 \rightarrow m3$ decay as an example again, the fitted lifetime is 5.00 ns. Going to the table we find two different decays from $m2$, one through $s4$ and one through $s5$, by now looking at the plot of the potential landscape again, we can realize that the one we want is the one through $s4$, and we find a calculated lifetime of 5.63 ns. The other fits are also reasonably close to the simulated results.

3.4 Discussion

The first part of this chapter was to some extent a sanity check. We validated that the model did not break down in the presence of thermal noise, and that the thermal behaviour was as expected. In this sense, there were not many new results in that section of the paper, it was rather elaborating on the findings presented in Chapter 2. One experimental prediction is that the effect of thermolubricity might be expected to be larger in atomically thin layered materials, due to the tip being able to slip to a lower band in the potential landscape.

However, in the remainder of the chapter where we discussed thermal relaxations we did make new experimentally verifiable statements. We predict that the thermal relaxation of an AFM tip on a sheet of layered materials will be much more complex than that on a crystalline material, and we give the times scales for this process. In particular, we expect thermal fluctuations to mask the true energy minimum state of the tip, and we instead propose that the tip will stochastically oscillate between multiple low energy states. This is not controversial from a PT-perspective, even in the standard PT model, the tip is expected to oscillate and visit the minima closest to the support position, but due to the more intricate two dimensional landscape, we expect this effect to be more pronounced here, and also that the tip as a consequence will spend more time away from the global minimum than in the PT model.

minimum	saddle point	$\Delta U/(k_\mathrm{B}T)$	lifetime (s)	minimum	saddle point	$\Delta U/(k_\mathrm{B}T)$	lifetime (s)
m1	s1	0.91	3.05×10^{-10}	m14	s22	15.02	5.54×10^{-05}
m2	s2	4.87	1.95×10^{-09}	m15	s20	8.87	8.95×10^{-07}
m2	s3	9.27	6.01×10^{-07}	m15	s23	20.28	1.18×10^{-02}
m3	s4	5.68	4.67×10^{-09}	m16	s19	15.06	4.83×10^{-04}
m3	s5	22.8	4.41×10^{-01}	m16	s23	25.85	3.15×10^{-00}
m4	s6	6.32	9.00×10^{-09}	m16	s24	26.21	4.63×10^{-00}
m5	s7	13.07	6.38×10^{-06}	m16	s29	22.92	2.33×10^{-00}
m5	s8	6.14	4.73×10^{-08}	m17	s24	20.08	9.44×10^{-03}
m6	s8	13.24	4.90×10^{-05}	m17	s28	15.13	4.90×10^{-04}
m6	s9	6.87	1.56×10^{-08}	m18	s25	14.83	4.52×10^{-05}
m7	s10	13.96	1.77×10^{-05}	m18	s27	7.10	1.34×10^{-07}
m7	s13	19.768	3.86×10^{-02}	m19	s29	0.05	5.87×10^{-10}
m8	s11	18.96	2.44×10^{-03}	m19	s30	20.92	3.20×10^{-02}
m8	s12	3.07	3.16×10^{-09}	m20	s28	2.34	1.76×10^{-09}
m9	s13	2.59	1.57×10^{-09}	m20	s30	24.87	9.34×10^{-01}
m9	s14	7.40	2.70×10^{-08}	m20	s31	27.25	$1.45 \times 10^{+01}$
m10	s12	15.07	3.83×10^{-04}	m21	s27	9.60	1.78×10^{-06}
m10	s15	14.55	3.36×10^{-05}	m21	s31	19.49	4.86×10^{-03}
m11	s16	19.76	6.64×10^{-03}	m22	s32	14.33	2.64×10^{-05}
m11	s21	15.82	9.36×10^{-04}	m22	s35	22.29	7.16×10^{-01}
m12	s16	26.76	8.36×10^{-00}	m23	s33	7.17	2.12×10^{-08}
m12	s17	25.34	1.75×10^{-00}	m23	s34	6.28	5.05×10^{-08}
m12	s20	6.60	1.01×10^{-07}	m24	s35	0.30	3.94×10^{-10}
m13	s17	20.59	1.65×10^{-02}	m25	s34	13.35	5.89×10^{-05}
m13	s18	36.00	$4.08 \times 10^{+04}$	m25	s36	13.64	1.25×10^{-05}
m13	s19	2.48	1.91×10^{-09}	m26	s37	6.64	1.25×10^{-08}
m14	s21	2.92	2.53×10^{-09}	m27	s38	0.06	6.81×10^{-11}
m28	s38	33.05	$4.17 \times 10^{+03}$				

Table 3.1: Tabuating all the minima from figure 3.6, complemented to each minimum is the available transition states (saddle points) to that minimum as well as the corresponding energy barrier, from this we can find the most likely decay path for the tip.

Part II

Collective Action in Social Networks

Chapter 4

The Dynamics of Social Networks

In this part of the thesis we will investigate the dynamics of attitude propagation in a social network. One could, and with good reason too, urge caution at this point, physicists making bold claims about describing systems far out of their field is something of a trope, and it does not always result in what is widely regarded as rigorous science in the recipient field, see e.g. theories of quantum consciousness [44, 45]. However, sometimes it does as in the observation of Bose-Einstein condensation in network theory [46], or the curious existence of the research field econophysics [47]. Keeping this cautiously in mind, we will go on to describe the concepts of agent-based models, social networks and network theory, and subsequently combine them with an Ising inspired model to obtain a model for attitude propagation in large scale common-pool resource games. On a note of form, this background chapter will be more thorough than previous ones. The reason for this is that this is a PhD thesis in physics, as such it is not assumed that most readers are already familiar with the fundamentals of social dynamics and network theory.

4.1 A trip down a peculiar path for a physicist?

We want to understand what drives cooperation in coordination problems. Perhaps the two most famous examples of such problems are The Prisoner's Dilemma [48] and The Tragedy of the Commons [49]. These two problems share a common origin, but they are qualitatively different. Consider the prisoner's dilemma:

Mr. Apple and Ms. Dash were caught roaming the streets of Celestia in the possession of dubious substances. They are now being interrogated by the authorities and are separately given the following deal: *We have strong reason to believe that you and your co-conspirator had plans to rob the Rarities Clothing store. We can put the both of you away for substance possession for six months. However, if you admit that the plan was to rob the store all along, then we are willing to drop the charges against you and instead settle for a two years sentence for the other prisoner. If the both of you admit the plan however, we will pursue a one year sentence for coming clean about it. Do you take the deal?*

The dilemma here is that – regardless of what the partner does – it is better for the individual to go for the plea deal, since they optimize for expected minimum time spent in jail. However, the collective minimum time spent in jail is found when no one agrees to the deal, that is to say, the most favourable collective state is reached when no one acts in their apparent self interest.

Now, consider the tragedy of the commons: *Five villages lie in the valley of a mountain. They all herd their sheep on the commons – the only greens found in the area. The sheep farmers have agreed to let their sheep graze the commons once a week, and that the commons are not to be grazed during the weekends to let the grasslands regenerate. One of the sheep herders decides to graze their sheep during Saturdays anyway, to produce a larger surplus of cheese in order to get an advantage over the others on market day. The other farmers follow suit one by one, and soon the commons succumb to over consumption and woe befalls all.*

The dilemma here is twofold: 1. Whether to stick to the consumption

quotas or not, and 2. How to deal with the situation of one herder not adhering to the agreement. Information in this system is not perfect, the herders do not know exactly where the limit for sustainable consumptions goes, but they do get continuous, albeit delayed, feedback from the vitality of the grasslands.

These are both examples of coordination problems. However, the approach to study them differ. The dominating approach to the prisoner's dilemma type problems is to use evolutionary game theory, in which a set of rational players will play to approach the Nash equilibrium [50], that is to say, the strategy that maximizes the probable personal profit of the individual players regardless of what action other players take. These are not the kind of games we will be investigating in this thesis, rather we will look into tragedy of the commons-type problems, where the players are prone to irrational decision making due to various psychological factors and the system complexity. In particular, we will develop an agent-based model to confront these issues. We will formally introduce agent-based models in a moment, but for now, think of them as models of groups of people where each individual is represented as a separate independent entity acting in accordance with some prescribed interaction.

A well-studied subset of the tragedy the commons type systems are *common-pool resource* games[1]. In this framework the agents can either be *cooperators* or *defectors* (terminology borrowed from game theory), depending on how they interact with the common-pool resource, i.e. the commons. It has famously been argued that the tragedy of the commons is an example of a pure cooperation game, that is to say, there inherently does not exists any technological solution to the game, instead, cooperation is the only resolution [51, 52], and in the event of sufficient defection, system wide tragedy might very well befall the inhabitants.

[1] A note on terminology: Different authors use different conventions here. We will associate the tragedy of the commons situations with common-pool resource games. But common-pool resource games are really just a subset of the more general so called common goods games, with the distinct addition of subtractability, meaning that in a common-pool resource game the resource consumption of one individual limits the consumption available to the next.

As a result, much interest has been shown to what drives cooperation and defection in common-pool resource games, for a formal and in depth analysis we refer to [53], where various cooperation strategies and their prerequisites are analyzed.

In this thesis we investigate *large scale* common-pool resource games. Where large signifies, large enough that individual agents cannot perceive the effect of their own, or other agents', individual resource consumption. The chief example of this would be consumption that directly, or indirectly, leads to the emission of greenhouse gases, where the commons then could be taken to be e.g. a carbon budget[2] [55]. Common-pool resource games of this scale introduce multiple obstacles to motivate agents to choose a cooperative strategy, or even realizing what a cooperative strategy might look like. In these kinds of games, we can find situations where seemingly rational strategies on the individual level cumulatively lead to a collapse on the collective level, because individuals are incentivized towards short term rational decisions, but which – obfuscated by multiple layers of complexity – ultimately results in devastating consequences on a global scale [56], and therefore by extension also for the oblivious individual. For an in depth deconstruction of these challenges within the scope of climate change, we refer to the analysis in [57].

In a game theoretical setting we would typically assume that all players are acting rationally. However, since it is hard for players to discern what a rational strategy might look like in the systems we are interested in, we will not be relying on classical game theory. We will rather mostly be relying on social dynamics – and opinion dynamics in particular – which describes how people interact and influence each others' behaviours and opinions to create collective habitual patterns. These models better capture social structures that might influence how people act, such as like-minded people sticking together[58, 59], how media influences a population [60], our confidence in other people [61], etc.

[2]For a general treatment of potentially cataclysmic global coordination problems, beyond common-pool resources, and the challenge of their mitigation, we refer to the excellent but harrowing account by Bostrom in [54].

Social dynamics is often modeled using agent-based models [62]. In an agent-based model, a network of agents with some connections between them is considered. Each agent can interact with any of its neighbours via some predetermined interaction, specifying the agents' means of expanding their influence in the network. Given some set of initial conditions (agent properties and network properties), the resulting dynamics can be investigated by running Monte Carlo type simulations where statistics are collected over multiple runs. In the language of network science, each unique network evolution sharing the same initial conditions is called a network realization. In order to be able to provide a more formal treatment of networks, we will introduce some network science terminology and algorithms in the following section.

4.2 The theory of (social) networks

A connected graph is a set of *nodes*[3], $N = \{n_1, n_2, \ldots, n_k\}$, connected by a corresponding set of *edges*, $E = \{e_1, e_2, \ldots, e_l\}$, where $e = e(n_i, n_j)$ designates the edge connecting the nodes n_i and n_j, forming a network where each node is connected to at least one other node via an edge. Furthermore, we call the total number of edges extending from a node the *degree* or *order* of the node, represented by the variable k. Every edge can be assigned a weight, w_i, which is typically taken to be a relative value, such that $w_i \in [0, 1]$, we denote the set of all edge weights as $W = \{w_1, w_2, \ldots, w_l\}$. Moreover we refer to a set of edges connecting two nodes as a *path*, and we define the *distance*, d, between two nodes to be the shortest path, or the geodesic, connecting them. In this formalism a network can abstractly be represented as a graph $G(N, E, W)$[4].

[3] In the literature a node is also referred to as a vertex in multiple accounts. The two can be used interchangeably, but "vertex" seems to be more common in graph theory, and "node" more prevalent in network science. Since what we do is network science, we will align ourselves with the conventions thereof.

[4] N.B. The traditional definition of a graph would be $G(N, E)$, and then introduce an associated weighting function mapping a weight to every element in E. Our definition is sufficient for our intended purposes however, because we will always

A graph can either be directed or undirected, depending on whether the individual edges are associated with a certain direction or not. We will however only concern ourselves with undirected graphs in this thesis, hence whenever we refer to a graph – or network – hereinafter, we will in fact be referring to an undirected graph. The average degree of a network is given by

$$\langle k \rangle = \frac{2|E|}{|N|}, \tag{4.1}$$

where $|\square|$ denotes cardinality i.e. the size of a set. Moreover, the average distance is given by

$$\langle d \rangle = \frac{1}{|N|(|N|-1)} \sum_{i \neq j} d(i,j). \tag{4.2}$$

Together 4.1 and 4.2 hold much qualitative information about the overall connectivity of a network. The connectivity of a network is intimately related to the topology of the network, and it is common to classify networks based on their global topology such as: A *lattice* is a network where every node is arranged on an ordered grid such that each node has the same number of neighbours (assuming periodic boundaries). The most simple example would be a square grid, but real-life examples would also include more elaborate topologies, such as hypercubic networking grids used in high performance computing clusters [63]. A *random* network is a network where every node is randomly associated with every other node according to some predefined distribution (usually Poisson) [64]. Random networks are a broad class of networks, out of which we are interested in a subtype called random *scale-free* networks. In these networks, the order of the nodes are distributed according to some power law, $P(k) \propto k^{-\gamma}$, where γ is some decay parameter describing how fast the degree distribution falls off.

have a well-defined, and predefined, set of weights to work with. The upside of this definition is that it simplifies the introduction of related concepts

A direct consequence of the degrees being power law distributed is that nodes of extremely high relative degree are bound to exist, which makes such networks exceedingly connected. The Internet is an example of a scale-free network [65]. Of particular interest to us, there are also *clustered scale-free networks*, these are scale-free networks, but with the additional structure that nodes tend to group up in collections, called clusters – or communities. A cluster is signified by a relative abundance of connections between the node members, as compared to non-members. We will return with a more formal treatment of clustering in a moment.

Social networks come in a plethora of sizes and topologies. For example, generically within a family, every single node (person) has a meaningful connection to every other node, on the contrary in a large corporation, as a rule, every single node does not have a meaningful connection to every other node. This raises the question of whether there exists a limit to how many individuals a person can maintain meaningful social relationships to. Dunbar famously argues that in primates this number scales with the size of the neocortex [66], and recent estimates put this number for humans somewhere in the range of 100-250 [67, 68]. In this thesis, we are however interested in modeling considerably larger networks than that. Large scale social networks, such as all user accounts on *Facebook* or *Twitter*; all the people in Sweden; or the most extreme case: every single person in the world, tend to naturally be organized in clustered scale-free networks [69, 70, 71, 72]. Naturally clustered because people tend to stick around in groups, e.g. families, fan clubs, fandoms, nations, political parties, companies, etc. The fact that these networks would be scale-free might at first seem counter intuitive. However, this is immaculately captured by Karithy's heuristic rule of six degrees of separation[5] [73], arguing that any one person is, by six social acquaintances, related to any other person in the world. This has become a well established fact which has been studied in multiple so called small-world experiments [74, 75], inferring that the grand network of humanity is highly

[5]Karithy initially muses this as a conundrum much more than a rule, but it has since been popularized as a rule.

connected. For the sake of scientific integrity, we should note that colleagues have recently questioned how common these networks actually are, see e.g. [76], criticizing among other things the rigidity of how the scale-free property is defined. We shall not be passing any judgment on this discourse, but rather stick to the established consensus that they are common, and use them as our prime representation for large social networks.

4.3 Reviewing the network construction algorithm

Having concluded that we will be using clustered scale-free networks as a basis for our simulations, we now turn to the question of how to obtain such a network. Fortunately, constructing – or growing – clustered scale-free networks is a fairly straightforward business. We will rely on the Holme-Kim clustered scale-free network growth algorithm [77] to accomplish this.

The Holme-Kim algorithm borrows from and expands upon previous results from Barabási et al. [78], and it involves three steps:

Preliminaries

Let $G(N, E, W)$ be a graph, and let E and W be empty such that no edges exist within the graph. N can either be empty or non-empty.

Growth step

Add a node, n_i, with degree, k_i, to G. For every edge of n_i, form a connection with some other node $n_j \in N$ with probability

$$P(e(n_i, n_j)) = \frac{k_j}{\sum_{n_m \in N} k_m}. \tag{4.3}$$

That is to say, the edge is formed with probability proportional to the degree of the considered node. This step ensures that the resulting graph is scale-free.

Clustering step

After a successful growth step have added a node n_i to the graph, a clustering step is initiated with probability

$$P_C = \bar{k}/(k_i - 1), \tag{4.4}$$

where k_i is the degree of n_i and $\bar{k}$ is a model parameter, corresponding to the average number of clustering steps to be taken per growth step. For the clustering step, let n_i and n_j be two nodes for which an edge was formed in the growth step, then add an additional edge between n_i and $n_k \in \Gamma(n_j)$, i.e. some neighbour of n_j. In the unlikely event that $\forall n_k \in \Gamma(n_j) \exists e(n_i, n_k) | e(n_i, n_k) \in E$, do an additional growth step instead.

Running this algorithm involves a tree of nested loops. During one master iteration step, the algorithm loops over N, and at every node visited, a growth step is performed. Then for every node added, the algorithm loops over all its principal edges (given by the order k), making connections as prescribed by eq. 4.3. Then the clustering step is initiated with a probability according to eq. 4.4 – if a clustering step is not initiated, then instead another growth step is performed. This process is then repeated until a sufficiently large network has been obtained. In the next section we will investigate how to actually characterize the clustering of these (and other) networks.

4.4 Communities and clustering

We have already established that communities are an important feature of social networks, people tend to gather in communities, and communities tend to organize in even larger communities of communities. Therefore we may surmise that communities are likely to impact how we approach large scale common-pool resource problems from an agent-based model angle. In network theory clustering means that given some graph, $G(N, E, W)$, G can be partitioned into a set of non overlapping subsets, which we refer to as *communities*, C. In order to systematically study communities we need to introduce a few more concepts

to our network theory toolbox. Using an approach similar to [79], let $\Gamma(n_i) = \{n_j | e(n_i, n_j) \in E\}$ be the neighbour list of the node n_i. Then we may can calculate the weighted degree of a node as

$$\tilde{k}_i = \sum_{n_j \in \Gamma(n_i)} w(n_i, n_j), \tag{4.5}$$

where $w(n_i, n_j)$, or w_k for short, is the corresponding weight to the edge $\mathrm{e}(n_i, n_j)$. Here we have also introduced a summation convention where the summation $\sum_{x \in S}$ indicates that the sum runs over all the elements x in the set S. The sum of all edge weights in the graph is

$$m = \sum_{w_i \in W} w_i, \tag{4.6}$$

Let $n_i \to C(n_j)$ be fuzzy notation for *the node n_i to the community $C(n_j)$ containing the node n_j*, as applicable in each individual case (see below). Assume that $C(n_i)$ is the community containing the node n_i, and let $W_{n_i \to C(n_i)}$ be the set of weights of edges connecting n_i to other nodes within $C(n_i)$ in accordance with our just introduced notation. Then we can calculate the sum of all edge weights of n_i inside $C(n_i)$ as

$$w_{n_i \to C(n_i)} = \sum_{w_k \in W_{n_i \to C(n_i)}} w_k. \tag{4.7}$$

Furthermore, we can calculate the sum of all the weighted node degrees in $C(n_i)$ as

$$k_C = \sum_{n_i \in C(n_i)} \tilde{k}_i, \tag{4.8}$$

we call this quantity the *community degree*.

Using these quantities we finally arrive at an expression for the *modularity* of a graph. Let $P = C_1, C_2, \ldots, C_k$ be the set of all communities given some partition of G, then the modularity, Q, of G is defined as

$$Q = \frac{1}{2m} \sum_{n_i \in N} w_{n_i \to C(n_i)} - \sum_{C \in P} \left(\frac{k_C}{2m} \right)^2 . \tag{4.9}$$

Modularity is another structural metric of a graph, beside the average degree and the average distance, which was previously introduced. It is a measure of the intra community connectivity, relative to the inter community connectivity. In this sense it is a measure of the quality (in terms of clustering) of a particular partition of a graph, a high modularity means that the nodes are strongly bound into communities. High modularity is a hallmark of many complex networks such neural and social networks [80]. In particular, we are interested in social networks, where clustering is a naturally occurring phenomenon, as such, we will want to ensure the networks we use have a high modularity.

Given some graph, $G(N, E, W)$, we can attempt to organize – or partition – it into communities. This requires a community detection algorithm. The aim of community detection is to find subsets of nodes that share as many strong (as measured by their edge weight) edges as possible. The potential number of subsets of a set grows exponentially with the cardinality of the set however. In the most general case where all subsets are valid, the number of subsets are given by $n = 2^{|S|}$, which computationally is an ominous growth rate to deal with. Fortunately, only a small portion of the possible subsets of nodes constitute valid communities in a network, since every node only has edges to a subset of the other nodes. Still, this exponential growth gives an idea about the challenge involved. One way of systematically doing community detection is by using modularity optimization. This approach gives us a quantitative way of ranking partitions of G by their associated modularity, the higher the modularity, the stronger the decomposition into communities, and hence, the better the clustering for our intended purposes. However, it has been shown that modularity optimization is a

non-trivial challenge in itself, in fact, it is an NP-complete problem[6] [82], meaning that its complexity is high, and in particular that it is not possible to reduce it into simple subproblems, each of which can be solved in polynomial time. To this end, the Louvian algorithm[7] was introduced by Blondel et al. [83].

The Louvain algorithm is a heuristic method for community detection using modularity optimization. It has seen great success in the field of community detection, and has, among other feats, in a serial implementation nonetheless, been used to successfully detect communities in a network consisting of 4 million cell phones with 100 million edges in bare minutes [84]. The Louvain algorithm operates on the ground of modularity gain, where a higher modularity is better. Assume that we have a graph, G, and that the node n_i belongs to the community $C(n_i)$. If we attempt to move n_i to the community $C(n_j)$ we can calculate the modularity gain by eq. 4.9 as

$$\Delta Q_{n_i \to C(n_j)} = \frac{w_{n_i \to C(n_j)} - w_{n_i \to C(n_i)\setminus\{n_i\}}}{m} + \frac{2\tilde{k}_i k_{C(n_i)\setminus\{n_i\}} - 2\tilde{k}_i k_{C(n_j)}}{4m^2}, \tag{4.10}$$

then the community assignment is n_i is decided by

$$\arg\max_{C(n_j)} \Delta Q_{n_i \to C(n_j)}, \quad \forall n_j \in \Gamma(n_i) \cup \{n_j\}. \tag{4.11}$$

Those expressions might seem a bit messy at first glance, but they are quite straightforward on a conceptual level. In the first equation we

[6] In order to not not make a long digression into computational complexity, we for this thesis note that a sufficient definition of an NP complete problem is: "A problem which time complexity cannot be described by a polynomial with finite coefficients and/or exponents.". In practice that means that the time it will take to solve the problem cannot be expressed by a simple polynomial [81].

[7] Trivia: This algorithm is in fact not named after one of its creators, or any other person for that matter, Louvain is in fact the French name of Leuven in Belgium, where the algorithm creators resided at the time.

calculate the change of modularity by comparing the sum of all edge weights in the community configuration prior to moving n_i to the sum after, and we add to this the difference of the corresponding sums for the weighted degrees of the communities. We then do this for every node in n_i's neighbour list, and we chose the move that generated the maximum modularity production.

Eqs. 4.10 and 4.11 provides our basis for comparing, evaluating, and acting based on changes in modularity of a graph. This leaves us to figure out how to choose $C(n_i)$ and $C(n_j)$, and also how to construct some initial partition of the graph to begin with. We outline how this is solved within the framework of the Louvain algorithm below:

Preliminaries

Let $G(N, E, W)$ be a graph. Fixate a modularity gain threshold, ΔQ_0, as a termination criterion for the algorithm. Introduce an arbitrary, but systematic, order of traversing all nodes in G. It has been observed that different orderings as a rule yield different final communities, however, the final modularity will not be notably impacted, although, the computation time can greatly vary [83].

Modularity optimization

Assume that every node is a single-node community such that: $\{n_1, n_2, \ldots, n_m\} = \{C_1, C_2, \ldots, C_m\}$. Looping over N, for every node, check its neighbour list $\Gamma(n_i)$, and find $\tilde{\Gamma}(n_i) = \{n_j \in \Gamma(n_i) | n_j \notin C(n_i)\}$, that is to say, all neighbours of n_i not in the same community as n_i. For $n_j \in \tilde{\Gamma}(n_i)$ calculate the speculative modularity change $\Delta Q_{n_i \to C(n_j)}$ should n_i change community to its neighour's community, $C(n_j)$, if that community has not already been considered for n_i. Do this for all neighbouring communities, and move n_i into the community corresponding to the highest modularity gain. In the case that all modularity differences are negative, the node stays in its present community. If a community becomes devoid of member nodes, it is nullified. When all nodes have been considered once, one iteration is finished, and the algorithm iteratively loops over all nodes again. This continues until no modularity gain is obtained by migrating any node to any other

community, whereupon the graph has reached a local modularity maximum, and the algorithm proceeds to the next phase.

Community aggregation

Coarse grain the graph by collapsing each community into a high level node, yielding a new graph $\hat{G}(N, E, W)$. These higher level nodes are taken to have an edge connecting back to themselves weighted by the sum of the intra communal edge weights of the underlying lower level nodes, i.e. $w(\hat{n_i}, \hat{n_i}) = \sum_{w_i \in W_{n_i \to C(n_i)}} w_i$. Conversely, let the edge weight of any two high level nodes be the sum of the edge weights of all inter communal edges between the corresponding lower level nodes, i.e. $w(\hat{n}_i, \hat{n}_j) = \sum_{w_i \in W_{n_i \to C(n_j)}} w_i$. Once the new coarse grained graph has been initialized following this protocol, repeat the modularity optimization step again using $\hat{G}$ as input.

Post processing

Iterate the first two phases until the modularity gain from community aggregation is below the specified threshold[8] ΔQ_0. Recursively unfold the top level nodes until the precursor nodes are retained, and let every node originating from the same top level node belong to the same community such that $C(n_i) = \hat{n}^{(j)}$, where $\hat{n}^{(j)} \supset \hat{n}^{(j-1)}, \ldots, \hat{n}^{(0)} \supset n_i$. This is the optimal partition of G into communities in terms of modularity.

In one summarizing sentence, what this algorithm does is that it identifies optimal communities of the nodes, and then identifies optimal mergings of communities, this is illustrated in figure 4.1. An alternative, and in some cases very valuable, interpretation of the second phase of the algorithm is that it identifies subcommunities to supercommunities, or, rephrased, it captures the hierarchy of the network. It is important to realize that the Louvain algorithm *does not* rewire the edges of a network at any instance. It simply sorts the nodes into abstract collections based on how the nodes are already arranged.

[8] For many networks it is viable to simply run the algorithm until it converges [83].

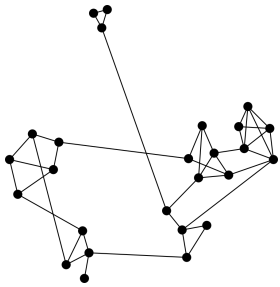
4.5 Some brief notes on implementation

Properly setting up networks requires a lot of work, fortunately the network science community is very active, and there are many free, open-source, high quality software packages available. In the work presented in this thesis, we have relied on the `networkx 2.3` package [85] for `Python3`. In this package, graphs are natively implemented as objects[9], with nodes being represented as dictionaries, and edges being neighbour lists implemented as dictionaries of dictionaries. For more information on the technical specifications, we refer the reader to the fairly elaborate package documentation[10].

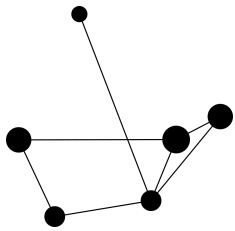
While `networkx` is an excellent tool for creating graphs in general, we especially want to optimize our network for clustering, since this makes for realistic representations of social networks. To this end, we employ the specialized graph package `community 1.0.0b1` in conjunction with `networkx`. This is a package that, among other things, implements the Louvain algorithm for community detection. We now have all the tools and theory needed to continue with setting up networks and studying their dynamics, this shall be the scope of the next chapter.

[9] Python is often regarded a scripting language, but it is also capable of object oriented programming when needed including C++-like classes, and more famously, dictionaries.

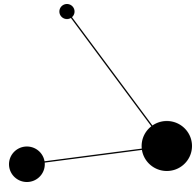
[10] Available here: `https://networkx.github.io/documentation/stable/index.html`.



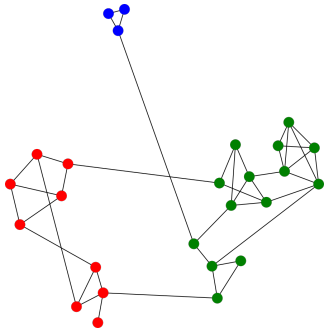
(a) First level nodes.



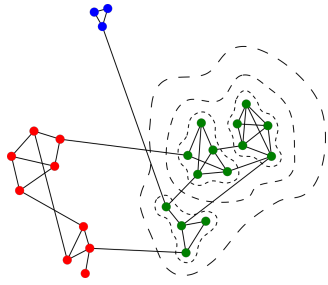
(b) Second level nodes.



(c) Top level nodes.



(d) Community classification.



(e) Showing community hierarchy.

Figure 4.1: Showcasing the Louvain algorithm for community detection. In this network, the edge weight is inversely proportional to the spatial separation of the nodes. We have omitted aggregate community self-edges for higher level nodes to increase the readability, c.f. *Community Aggregation* in the algorithm description.

Chapter 5

Attitude Propagation in Social Networks

In the following chapter we present our model for opinion dynamics in relations to a large scale common-pool resource. We adapt an Ising inspired statistical physics approach to obtain an agent-based model capturing real-life features such as echo-chambers and late stage opinion homogenization. We investigate how the model behaves on various networks and make predictions about which factors can speed up the convergence of these networks, and to what degree. This chapter summarizes Paper III.

5.1 Modeling social dynamics

Quite unsurprisingly, modeling the dynamic spread of ideas in complex social networks turns out to be hard. It is, in essence, a *messy* problem, this can be argued from a simple thought experiment that can be performed by anyone partaking on modern society at any time, ask yourself: *How many times in the last week have I been prompted to update my opinion on climate change / the US president / Internet privacy*

(or whatever other polarizing issue you engage in)? And through what channels have I been influenced? Most likely, you have not changed your view noteworthily on any issue that you care strongly about (we will return to that in a moment), but you would likely have been exposed to a multitude of people or institutions trying to reinforce or change your opinion. You talked to a friend, you read the news, you overheard a conversation in the coffee room, you saw an advert, *et cetera*. Modeling all of the available pathways, and how opinion spread through these, would be nigh impossible. Fortunately, in this thesis we aspire to contrive *simple models*, for that we do not need this level of detail. Using a simple model we can then get a qualitative idea about how ideas spread.

Crucially, we will start by dividing the mechanisms that propagate opinions into two overarching categories: I. Local effects: mechanisms that act on an agent-to-agent level, and II. Global effects: mechanisms that act on all agents at all times. In terms of these categories, we can characterize mechanisms impacting the transfer of opinions in terms of the following subtypes:

Type I a) Mechanisms that influence how susceptible an agent is to any other particular agent's influence, e.g. friendship or animosity, a person's ethos, if you speak the same language.

Type I b) Traits that work to preserve an agent's held beliefs, e.g. their conviction or stubbornness.

Type I c) Stochastic local external factors, maybe the agent was in a foul mood at the moment the interaction occurred, or maybe the agent had just read a particularly touching article, maybe they were interrupted.

Type II a) Channels which effects are shared by all agents, such as relevant policies, media reporting, the moral hegemony of society.

Knowing what parts we want in our model we can now set out to actually construct it. Let opinion be a real valued variable $a \in [-1, 1]$,

we shall commonly also refer to this as the *state* of an agent, in a bit of terminology borrowed from physics. Let the result of an interaction between two agents (labeled by i and j) be

$$\Delta a_i = |a_i - a_j|[w^-(1 - a_i) - w^+(1 + a_i)], \tag{5.1}$$

where w^- and w^+ are the weights for an agent becoming more, or less, cooperative respectively. We have $w^- = 1 - w^+$, and we can introduce a function $f(x)$ as below to calculate these weights using $w^+ = f(x_{ij} + \xi)$. We can simplify equation 5.1 slightly and obtain

$$\Delta a_i = |a_i - a_j|[2f(x_{ij} + \xi) - 1 - a_i] \tag{5.2}$$

where,

$$f(x) = \begin{cases} 0 & \text{if } x < -r \\ \frac{1}{2r}(r + \xi) & \text{if } x \in [-r, r] \\ 1 & \text{if } x > r \end{cases} \tag{5.3}$$

and,

$$x_{ij} = w_i a_i + w_{ij} a_j + \phi. \tag{5.4}$$

where we have introduced some parameters to contain the influence modifiers discussed above. The parameters w_i and w_{ij} are the self weight, and the weight between agents i and j, respectively, this represents type I a) and I b) mechanisms above. Furthermore ξ is a uniform random noise term, distributed over the interval $\pm r$, this is a representation of type I c) mechanisms. Finally ϕ captures type II effects, it is a field-like parameter, and at times we shall refer to it as *the external field*[1].

[1] If this sounds like physics, it is definitely a testament to the origins of this interaction. Upon closer inspection the shrewd reader might see similarities between this model and the interaction as formulated in the Ising model for paramagnetism. This is indeed the case, this model was inspired by the Ising model.

The first factor in equation 5.2 is there to limit the effect two already agreeing agents have on each other. Groups of agreeing agents clustering together in so-called echo-chambers to reinforce the internal opinions of the group is a real effect. However it is not something that primarily occurs on an agent-to-agent basis—it is a mesoscopic effect founded in the local network topology, as we shall establish in a moment.

These parameters are subject to a set of standard values that will be used throughout this thesis if nothing else is explicitly stated, they follow here. The external field is taken to be $\phi = 0.05$, there is no clear rationale to this number in particular, rather it is chosen in relation to all the other parameter values to yield reasonable convergences, however, it should at least be positive to reflect a society that incentives being a cooperator. The agent-to-agent weights, w_{ij}, are picked from a normal distribution with $\mu = 0.5$ and $\sigma = 0.15$, where μ and σ are the conventional mean and standard deviation. These were chosen with regards to the self weight, which was taken to be $w_i = 0.6$, for comparison allowed values for the weight are 0 to 1. We assume that people are generally more prone to keep their own opinion than they are to change it, and furthermore that they are quite prone to do so, hence these numbers in specific. The random noise, ξ, is taken to be uniformly distributed rather than normally distributed. The reason for this being that it is in principle a white-noise term to the interaction, so it would be uncorrelated and uniform. The width of this noise is set to $r = 0.1$, the exact value of this number is not expected to be crucial beyond being small but significant. These parameter values are chosen such that every network realization should in the end become fully cooperative. We will come back to why that is in a moment.

That is all for the parameter values. However, there are other initial conditions that need to be addressed as well. We have already established that we will be using clustered scale-free networks as our network topology of choice. The size of the networks we settled on was $33^2 = 1089$, this quirky value in specific because it happened to be the first square number larger than 1000 that came to mind at the time[2]. It had to be

[2]Yes, whatever happened to 1024?! Alas, now we are stuck with this number.

a square number to enable comparisons to square grid networks. One may reasonably discuss exactly how large a network has to be in order to qualify to model "large scale" common-pool resource games. To obtain qualitative results, having a huge network is however not as important as it might at first seem. Crucially, a network needs to be large enough to house the chosen network topology. In our case that means the network needs to be large enough that: there should be tangible clustering and there should be a scale-free relationship between those clusters. We will return with a remark on finite size effects later on. In relation to generating the correct topology, the average degree of the network was set to $\langle k \rangle = 8$. Take note that this is the *average* degree, as outlined in the last chapter, the actual degrees will be distributed according to a power law.

Initially, the states of the agents are distributed by a gaussian with $\mu = -0.25$ and $\sigma = 0.15$. This is chosen with regards to modern western society, where we, on a collective level, are prone to massive over consumption such that most people are – knowingly or unknowingly – operating under a defecting strategy [86, 87]. This being a consequence following the lack of clear feedback loops in relation to the resource consumption, as outlined in the previous chapter.

The algorithm that simulates the opinion spread in networks based on the interaction above is exceedingly simple. It iterates four steps:

1. Pick a random agent.
2. Pick an agent from that agent's neighbour list.
3. Let the agents interact.
4. Update the first agent's state.

An important detail to point out here is that in step four, it is the agent that was singled out in the first step whose state is updated. This ensures that agents with more connections are more likely to spread their opinions, which would seem realistic. Finally a quick note on presentation. As just stated, only one interaction occurs every timestep, which would seem slightly artificial, therefore we will be introducing a more

organic time unit which is [time] $= t/N$, where t and N are the elapsed time in number of timesteps, and N is the system size. Meaning the unit of time is the time it would take for every agent in the network to make one interaction each.

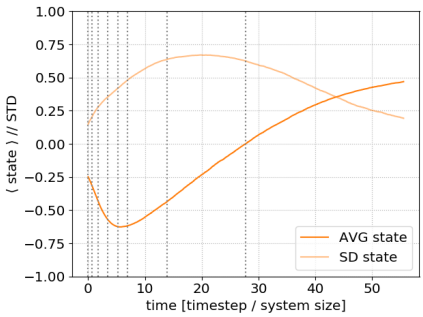
This is our simple model for opinion propagation in its entirety. The next step is to construct a series of networks and see what kind of dynamics it produces. In principle, nothing in the model critically relies on a specific network type – i.e. qualitatively the dynamics will be the same regardless of network type within reason. Realizing this we will start off by studying the model on a simple grid network, and then transition our findings into complex clustered scale-free networks, this should improve our understanding of how the model functions.

5.2 Interpreting model data

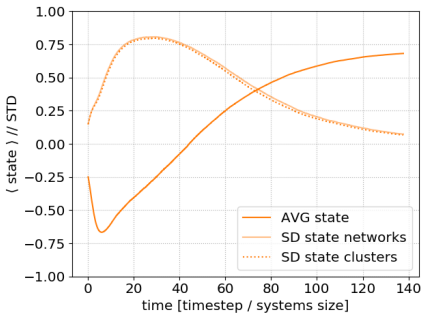
When we look at grid networks, we will initially study smaller networks ($12^2 = 144$ nodes with periodic boundaries), because they are easier to visually inspect. In figures 5.1a and 5.1c-j, the evolution of the average state is given along with some corresponding snapshots – in the snapshots blue corresponds to defecting, and red to cooperating. It is useful already at this state to get familiarized with the fact that the average state is generated from hundreds (500 to be exact) of realizations of the network. Hence, one should be mindful when making claims about individual realizations from the average and vice versa.

Comparing figure 5.1a with 5.1c-j, we can see how initially the average state decreases, up until some point, where it instead starts to increase, and finally it looks to possibly be converging to some value. The standard deviation follows a similar but opposite sequence, it initially increases but after some time it decays away. Let us spend some time discussing what we see here, before we transition into more complex networks.

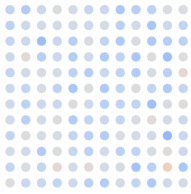
Initially, the network is slightly defecting and largely unpolarized – po-



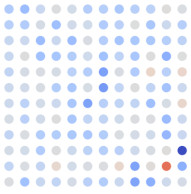
(a) Network evolution on a grid, with snapshots indicated.



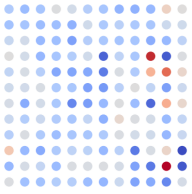
(b) Network evolution on a clustered scale-free network.



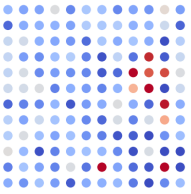
(c) t = 0



(d) t = 100

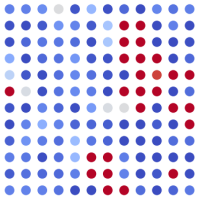


(e) t = 250

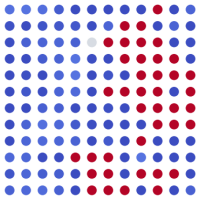


(f) t = 500

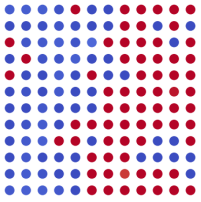
(g) t = 750



(h) t = 1000



(i) t = 2000



(j) t = 4000

Figure 5.1: (a) An example of how a grid network might evolve, in this example most of the networks become cooperative. (b) A typical clustered scale-free network evolution with parameters slightly favouring cooperation. (c-j) Snapshots of one realization corresponding to (a). We can see how the networks initially rapidly polarizes, and then slowly converges.

larization here being some measure of to which degree the agents in the network share a common opinion[3]. As the network evolves we see two things happening in parallel: 1. Some agents are *radicalized* and start to populate their surroundings with their own opinion (e.g. node 10,4 indexed from top left). 2. Overall the network is gradually becoming increasingly defecting. Let us start by tracing these two observations back to their origin in order to understand how they come to be. The first case, of strong opinion agents converting their neighbours, is a direct result of the interaction, eq. 5.2. There are two reasons for this, firstly the prefactor allows for a large Δa_i when the agents' opinions differ significantly, and secondly the interaction term x_{ij} puts a large emphasis on having a strong opinion. This gives rise to *clusters*[4] of agreeing agents. Meanwhile, the rest of the network will initially grow increasingly defecting, since on average, we will find two already defecting agents interacting, which is expected to slightly increase the level of defection. In the event that two agents of equally strong but opposed opinions interact, the external field will, on average[5], push the interaction in favour of cooperation (since we stipulated $\phi > 0$).

Following this development clusters of defecting and cooperating agents will emerge. We call this the *cluster formation regime*, and it continues until there are no more unpolarized regions left in the network. Take note that the end of the cluster formation regime does not coincide with the moment of minimum average state, the former is rather characterized by the standard deviation saturating. When cluster formation has concluded and the network is fully polarized, then any meaningful interaction is taking place only at the boundaries of disagreeing clus-

[3]We intentionally use polarization in an open-ended manner here. Quantitatively, polarization is to a large extent coupled to the standard deviation of the average state, however, on a conceptual level we rather use it as an inverse measure of how many neutral agents remain in the network.

[4]N.B. clusters are not the same as the communities introduced in the previous chapter – communities are topological collections of nodes, clusters are nodes sharing an opinion. Initially these will be correlated, but clusters will quickly outgrow their communities.

[5]In any one specific interaction the random noise will play a significant role, however, its effect averages out over many interactions.

ters. Inside of clusters any interaction will only serve to preserve the members' opinions, this is a representation of echo-chambers, which are groups characterized by opinion homogenization. An appealing point to this model is that echo-chambers are not introduced *a priori* on an agent-to-agent basis—they are a mesoscopic emergent topological phenomenon arising from several interconnected agents sharing a common opinion, just like we would expect from reality.

We refer to this second regime as the *equilibration regime*. Since any meaningful interaction at this point will be between agents of extreme opposite opinions, the external field plays a pivotal role here. There are only two possible futures available to the networks: either become fully cooperative, or fully defective. When a network becomes fully homogenized to either state, it is no longer possible to change the state of any agent, hence this is an absorbing boundary of the system. When switching to a clustered scale-free network, the evolution is qualitatively the same. We have given the evolution of a clustered scale-free network under similar conditions, but with 1089 agents, in figure 5.1b. The standard deviation here is the average of the standard deviations in each network (or their clusters).

It is easy enough to imagine that the terminal state is largely dependent on the relative abundance of defectors at the time of saturation versus the strength of the external field. We will spend some time elaborating on what other factors play a role here, but chiefly, we will be investigating how these factors impact the convergence time rather than the convergence state. The reason for this is that quite some attention already has been given to the issue of finding *the tipping point* of these systems [88, 89, 90], that is to say, how the terminal state depends on the system parameters. However, little attention has been given to the actual convergence time. While the terminal state of the networks is of obvious principal interest, it is, in fact, the factors that influence the convergence speed that is of immediate practical interest. If we want to answer questions such as: "What interventions would be effective ways of mitigating the climate crisis?", then yes, one part of that answer is to find *an* intervention, but the real challenge is in finding *an effective*

intervention. The reason for this being that we might not have the time to wait for an inefficient action to come into effect, we need to act swiftly to avert a potential global catastrophe. This is the rationale for choosing parameters that ensure that the system ends up fully cooperative, it is really the only acceptable end result, and we are studying how to get there as fast as possible.

Having identified what dynamics drive the evolution in grid networks, we will now be switching to the larger and more realistic clustered scale-free (hereinafter: CSF) networks. Without further ado, we provide a comparison between clustered scale-free networks and grid networks (this time at 1089 agents) in figure 5.2. As expected we see a similar network evolution as before. The CSF evolution is however faster than that of a grid network. We reason that this is because in a CSF network the topology is more complex in such a way that the boundary-to-area of clusters is higher. That is to say, there are more boundary interactions in a CSF network, and since those are the only meaningful interactions after cluster formation due to echo-chamber effects, these networks will converge faster. Or, equally true: it is harder for clusters to protect their members because more members are exposed on the boundary.

We already remarked that it is hard to visualize CSF networks, however, at this point it is warranted to at least show an example of a small scale CSF to see what we are dealing with, therefore, a snapshot of a CSF network on the path to cooperation is given in figure 5.3. The left hand side of the plot displays the nodes and their state, with the size of the nodes proportional to their degrees. The right hand side is displaying clusters as identified by the Louvain algorithm, as well as various quantitative information about the clusters. This image shows that the same clustering is present in CSF networks as in grid networks, albeit in a more complex representation. This network is just an example, and it does not reflect any actual scientific results presented in this thesis.

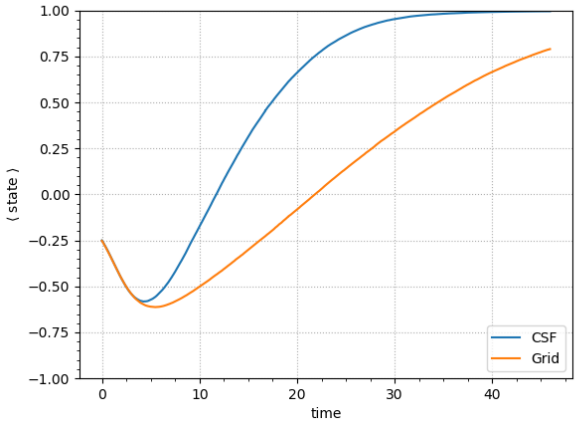


Figure 5.2: Investigating the quantitative change when going from a grid network to a clustered scale-free (CSF) network. We see how the convergence speed increases when we change the network type, this is due to the increased complexity of the CSF network.

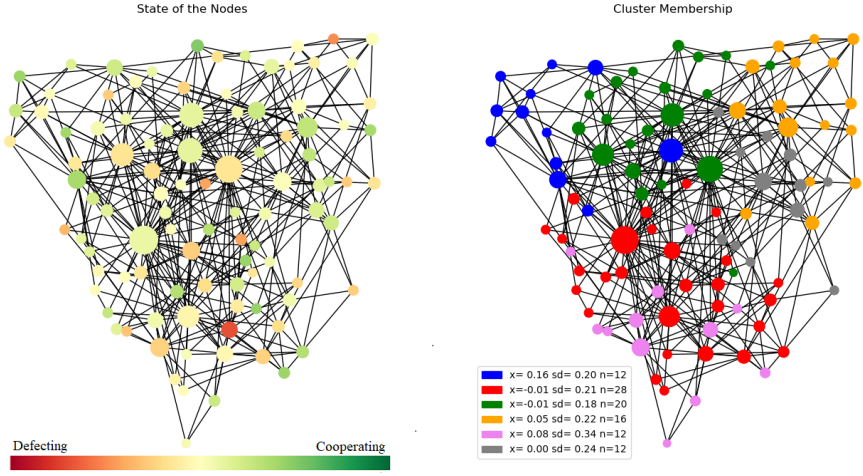


Figure 5.3: An example of clustering as identified by the Louvain algorithm on a clustered scale-free network as created by the Holme-Kim algorithm. Image adapted from [91].

5.3 Convergence rates in social networks

Our next step is to start investigating the effect of individual model parameters on the convergence rate. However, in order to do so, we need a consistent definition of convergence for our systems. We observe that the equilibration regime seems to be characterized by an exponential decay of the form

$$\langle S \rangle = 1 - e^{-\gamma t}. \tag{5.5}$$

We can rewrite this expression to isolate the rate parameter γ as

$$\frac{d\langle S \rangle}{dt} = \gamma e^{-\gamma t} \quad \Longrightarrow \quad \gamma = -\frac{\langle \dot{S} \rangle}{\langle S \rangle - 1}, \tag{5.6}$$

where $\dot{\square}$ conventionally denotes the derivative with respect to time.

In order to ensure that the data sets become comparable, it is important that the above derivative is calculated in a consistent way. We will simply always evaluate it at the point closest to zero average opinion. The derivative itself can readily be found by linear regression, we applied a 10 point regression window to safely eliminate possible numerical instabilities.

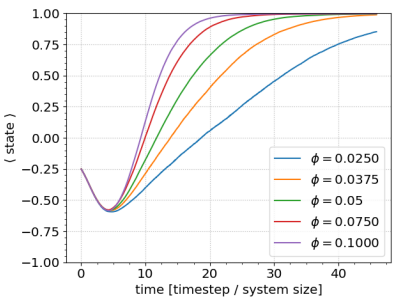
We have already established the importance of the external field for driving the networks into cooperation, it shall hence be our first case study when investigating the convergence rate's parameter dependence. This prompts us to pick a point at which to change the field. Two alternatives would be to either use a different external field from the beginning, or delay changing it for some time. The better alternative of the two is to change it at a later time, the reason for this is twofold: 1. Immediately changing the field will to a large extent change the ratio of cooperators and defectors at the end of the cluster formation regime, and subsequently the effect of this will be felt throughout the

equilibration regime, which leads us to 2. Changing the field at a delayed point isolates the effect of the field on the convergence.

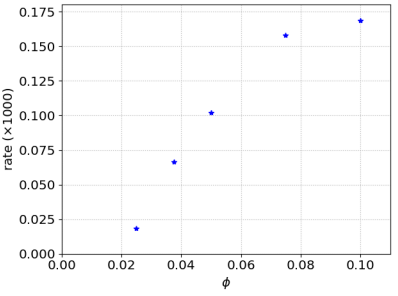
Ideally, we would make the change in the field at the point of maximum polarization, which is where the equilibration regime begins. However, it is hard to pinpoint an exact point where this transition occurs, because, as can be seen in the standard deviation part of figure 5.1b, there is a fairly large plateau somewhere along which the actual point of maximum polarization happens. Moreover, the average state changes significantly over this interval, so there is room for errors. Hence, we have instead elected to use the point of minimum average state as the point where we change the external field, since it is more well defined. This is possibly also an appealing point to choose from a real world perspective. It seems plausible that the external field will be reactive to the average state – when there is a problematic level of defection, regulations will be enacted to incentivize the agents into cooperation. From this reasoning it is not unreasonable that the lowest average state would be a starting point for increasing the external field.

Keeping this in mind we give the external field as well as the corresponding rate dependence in figure 5.4. As expected, the convergence rate follows the strength of the external field nicely. For significantly low fields (outside what is displayed here), we will find cases where the average state does not converge to 1. These cases are briefly investigated in Paper III, and have already been thoroughly investigated by colleagues looking into social tipping points, hence we will leave them out here. We see a nonlinear response in the rate yield after some point beyond $\phi > 0.075$. This is because we are approaching the limit where the field overtakes the noise intensity, meaning that we are closing in on the maximum convergence rate.

Looking at the model interaction, eq. 5.2, we realize that the only realistic control parameter we have is really the external field, as such, we may take the approach above as exhaustive as far as tweaking the interaction parameters goes. This leaves us with making changes to the network itself as our next option. We already to some extent studied the connectivity of the network when going from a grid to a CSF network.



(a) Some different external fields

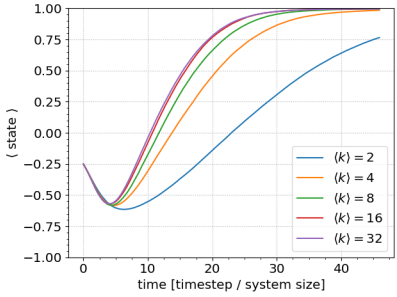


(b) The corresponding convergence rates

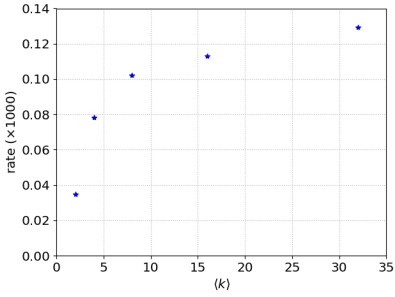
Figure 5.4: Changing the external field and investigating the effect on the convergence of the average state within a factor of two. We observe how it has a strong effect on the convergence speed.

We can make this investigation more consistent by keeping the same network type (CSF), and instead change the average degree of the network. Such a comparison is provided in figure 5.5, with an accompanying rate plot. We see how the convergence speed indeed increases with improved network connectivity. The reason for this being the same as in the grid network – it gives the dominating faction better circumstances for dissolving opposing clusters. Notwithstanding, the rate gain drops sharply for very high connectivities. This is due to finite size effects. Any given agent has on average $\langle k \rangle$ nearest neighbours, it follows then that they have on average $\langle k \rangle^2$ next to nearest neighbours. If we calculate the number of next to nearest neighbours for each network connectivity we end up with $4, 16, 64, 256, 1024$, or relative to the total number of agents $0.3\%, 1.4\%, 5.9\%, 23.5\%, 94\%$, certainly we would experience finite size effects at least in the last of these cases.

We can cement this statement by taking a moment to visualize how a network would look where the distance from any agent to any other agent was 2. Recall that any meaningful interaction in these networks happens on the boundary between two clusters. Then, if a boundary is taken to be one agent "thick", and the distance between any two agents



(a) The average state response to changing the average degree.



(b) The convergence rate response to changing the average degree.

Figure 5.5: Investigating how the average degree impacts the average state and its convergence rate as it is changed by multiples of two. For high enough connectivities, finite size effects will limit the effect of increasing the connectivity

is two, the boundary-to-clusters ratio is going to be huge. It follows from this that it will be very hard to form any meaningful clusters at all to protect members from outside influence. Hence going from say, an average degree of 30 to 32 does not really matter, because it is not going to influence the rate at which clusters are dissolved noteworthily, as compared to going from an average degree of 2 to 4, which increases the amount of boundary significantly.

Recapping, so far we have tried changing what parameters we realistically are in control over and we have tried changing the connectivity of the network. These are both low hanging fruits, now it is time to consider what outside the box ideas – within our sphere of influence – we can identify to significantly impact the convergence rate. If we take a contextual approach and draw some inspiration from the present-day digital social media networks, the 2010s have made it abundantly clear that so-called *infuencers* have massive social traction [92, 93]. In our model, an influencer is basically an agent of extraordinarily high degree. Such agents are bound to exist, since the node degrees are power law distributed, as is required for the network to be scale-free. We could

then play the following shenanigans: First we let the system polarize in a similar way as we did when examining the external field, then we identify the agent of highest degree and swap this agent for an unyielding cooperator and watch how the network evolves.

In this way, we can introduce a benevolent influencer to the network. However, this scheme potentially does two things. As intended it places a cooperator in an especially influential position, but it also introduces an agent whose state cannot be changed. We should make sure that the effects we see are actually coming from the agent's local network topology, not simply the fact that the agent has a fixed opinion. We accomplish this by comparing the effects of adding an influencer to a CSF network with adding one to a grid network, since a grid network is unbiased as far as agent degrees goes. The result of doing such a comparison is displayed in figure 5.6. We see how this intervention has almost no effect in a grid network, whereas the effect on a CSF network is palpable. This makes a strong case for the fact that connectivity is a key feature to increase the convergence rate in these networks.

To get an idea about the relative rate increase of various interventions, we gather the strongest contenders in one plot, given in figure 5.7. We see how the strongest effect is obtained by changing the external field by a large amount. The reason this is outweighing increasing the average degree is that even if the connectivity of the network is large, the amplitude of any given Δa_i is still limited by the external field (in the event of maximally polarized counterparts). Quite surprisingly however, the result of drastically increasing the average degree of the network, matches the impact of adding *one* influencer. This speaks to the profound importance highly connected agents play for the network.

5.4 The future, odds & ends

One point of scrutiny should be examined, in order not to leave some inquisitive readers sceptical: "Why would a self-proclaimed model about common-pool resources not make any reference to a common-pool re-

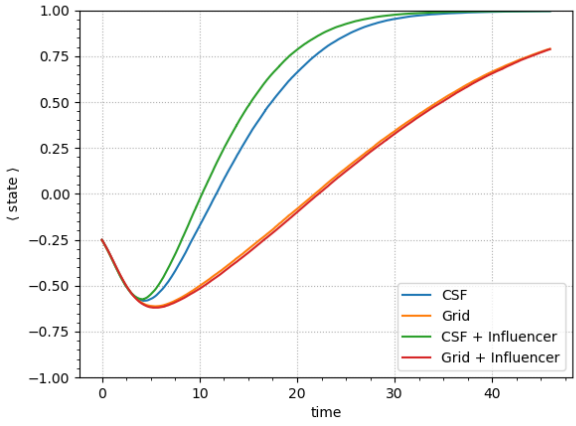


Figure 5.6: Adding an influencer to the system and comparing its effects to that in a grid network. We see how the effect is tangiable for CSF networks.

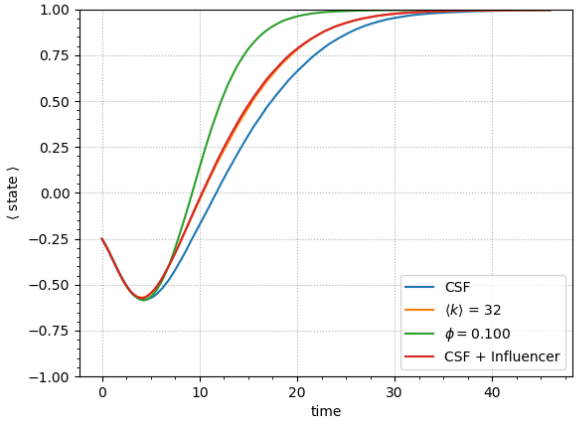


Figure 5.7: Comparing the various interventions we have studied to improve the convergence rate. In the order of the plot legend:

source?". This is a fair interjection, we set out to model some common-pool resource, we even mentioned already that one such resource might be consumption of the collective carbon budget. Yet, in the model there is no mention of any such resource at all. The reason for this is really provided in the previous chapter, we define a large scale common-pool resource game as a scenario where agents cannot see the effect their consumption has on the common-pool resource i.e. the resource appears to be unlimited – or conversely the consumption appears infinitely small. The point here being that there is no tangible feedback mechanism for the agents to relate to.

Typically, in the common-pool resource game literature, one would talk about the defector utility, that is to say how much an agent gains from choosing a defecting strategy, and weight that against the ramifications thereof, probably analyzed through a causal loop diagram to visualize the underlying feedback loop. In the absence of clear ramifications however, it is not immediately clear how to capture it in this framework. For better or worse, our approach has been to circumvent the feedback loop altogether and instead make an agent-based model to make qualitative predictions about how people might react in this common-pool resource problem. Better, because in this way we have been able to construct a quite general model that operates on a large scale, unifying both the clustering and equilibration regime in one picture, while also giving qualitative insights into how to create an environment that favours cooperativity in a societal collective action issues in a broader context. Possibly worse because feedback loops make for simple verification of a model and also gives clear directions on how to proceed with changing people's strategies.

An interesting account that might inform further development along the lines of actually introducing a feedback loop into these systems is provided in [94], where the authors discuss social ostracism as a "resource". Here a level social ostracism is attached to the consumption of an otherwise abstract resource. This ostracism would be tangible, swift, and reactive, all which make for a good feedback loop.

Going forward, it would be interesting to see how this model relates to

empirical data. There is empirical evidence, such as [95], which give support to so called bounded confidence models (elaborated upon in Paper III). This also supports the cluster formation dynamics observed within this thesis. There are also empirical evidence supporting consensus as a long term state [96]. Going forward, it would be interesting to study specifically the transition between these two regimes. Furthermore, it would be valuable to empirically evaluate the interventions suggested herein. Due to the availability and size of the Twitter data set, a first step could be to investigate how influencers on Twitter homogenizes the opinions of their followers.

Finally, we would like to point out that this model can be used to describe a wide range of large scale common-pool resources. Nothing we have introduced limits the model to climate change, we have used carbon budget consumption as an example to have something real to relate our results too.

Part III

Stochastic Thermodynamics and Image Reconstructions

Chapter 6

Free Energy Image Reconstruction Processes

In this chapter we will introduce three different image reconstruction protocols and compare their performances. However, in order to understand the workings of these algorithms, we will start by introducing the so called Jarzynski equality, and some accompanying stochastic thermodynamics. Two out of the image reconstruction algorithms will be binning based methods: the weighted histogram method, and the explicit current method, and one will be a traditional noise reduction algorithm: the Lucy-Richardson deconvolution. This is still a work in progress, and it is presented as is. This chapter is attached to no paper for this reason.

6.1 Microscopic stochastic thermodynamics

Generally, macroscopic thermodynamic systems in, or close to, equilibrium is governed by the influential *Laws of Thermodynamics*. In this part of the thesis, our end-game is to reconstruct free energy surfaces from AFM experiments. Those experiments are neither macroscopic nor are they performed at equilibrium. To achieve this goal we would like

to have similar laws, pertaining to these new conditions. That is the purpose of this section.

Consider a system, coupled to a heat reservoir of temperature T, and with a corresponding Hamiltonian $\mathcal{H}(z, \lambda)$, where z is a dynamical variable and λ is some externally tuned parameter driving the system out of equilibrium, such as an electromagnetic field. Furthermore assume that the system is in equilibrium. If we then change λ according to some function, or protocol, $\lambda = \lambda(t)$, for some finite time, τ, we will drive the system out of equilibrium. This will trace a trajectory, Γ, in the phase space of z. Since the dynamics of thermodynamic systems are generally stochastic in nature, Γ will tend to be non-deterministic, hence a unique trajectory will be obtained every time the system is driven out of equilibrium. In particular, for our AFM experiments, the dynamical variable will be the position of the tip, the external parameter is the position of the support, and our protocol will be to move the support at constant speed. Meaning, we will use $z(t) = x(t)$ and $\lambda(t) = vt$ from now on.

The change in energy in our system can be written

$$dE = -\frac{d\mathcal{H}(x, \lambda)}{d\lambda} d\lambda + \frac{d\mathcal{H}(x, \lambda)}{dx} \circ dx, \tag{6.1}$$

In this expression, the first term is the change in energy due to the system being driven by λ, i.e. the work done on the system, dW. Furthermore the second term is the change in energy due to configurational changes in the system, corresponding to a change in entropy dS, the second term thus becomes the dissipative heat $dQ = TdS$. Moreover, we have introduced $\circ$ in response to the fact that the x-coordinate is subject to stochastic noise, and we shall have to treat it using Stratonovich integration. In this framework we use a midpoint approximation to integrate an arbitrary function along a stochastic trajectory, as

$$\mathbf{f}(x) \circ d\mathbf{x} = \sum_i \mathbf{f}(\frac{\mathbf{x}_i + \mathbf{x}_{i+1}}{2})(\mathbf{x}_i - \mathbf{x}_{i+1}) \tag{6.2}$$

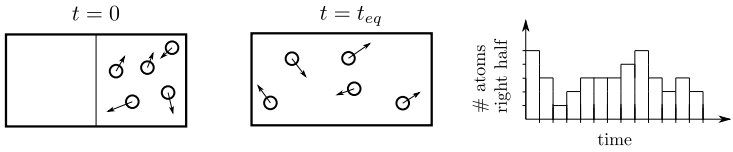


Figure 6.1: Microscopic entropy fluctuations visualized using the position of some atoms in a box. As the system evolves, there is always a small probability of finding the system back in its initial state with all particles in one half of the box.

Integrating eq. 6.1 over a phase space trajectory, we obtain expressions for the work and dissipative heat on the microscopic level [97, 98]

$$W[x] = \int_0^\tau \frac{\partial \mathcal{H}}{\partial \lambda} \dot{\lambda} dt, \tag{6.3}$$

$$Q[x] = \int_0^\tau \frac{\partial \mathcal{H}}{\partial x} \circ \dot{x} dt. \tag{6.4}$$

where we have introduced $[\square]$ to denote path integrals (in the phase space).

Next we turn to the concept of entropy. On the microscopic level, the probability of going from a higher to a lower entropy state will not be negligible, c.f. fig. 6.1.

This is captured by so called fluctuation theorems [99, 100, 101]. There are many such theorems, the first one to be formulated however was the transient fluctuation theorem [99], which states

$$\frac{P[\Gamma(t)]}{P[\hat{\Gamma}(t)]} = e^{S[\Gamma(t)]}, \tag{6.5}$$

where $P[\Gamma(t)]$ is the probability of the trajectory $\Gamma(t)$, $S[\Gamma(t)]$ is the entropy production along $\Gamma(t)$, and $\hat{\square}$ denotes the reverse direction in phase space.

This expression tells us that small negative changes in entropy, while staunchly suppressed, is feasible. Conversely transitions to higher entropy states are exponentially more probable. This is then our microscopic analogue to the second law of thermodynamics.

Crooks provides the following fluctuation theorem for a system starting in equilibrium [102]

$$\frac{P[\beta W[\Gamma]]}{P[\beta \hat{W}[\Gamma]]} = e^{\beta(W[\Gamma]-\Delta F)}, \tag{6.6}$$

where $W[\Gamma]$ is the work done along the trajectory Γ, ΔF is the difference in free energy between the initial and terminal states, and $W[\Gamma] - \Delta F$ is the corresponding dissipative work.

This theorem captures the probability of different trajectories starting from equilibrium in terms of their corresponding works. In the AFM experiments we will be studying the tip will start out in some potential minimum, and then be pulled out of equilibrium by the moving support, so this is starting to look like something that could be of use to us.

Crooks later generalizes this fluctuation theorem in reference [103], stating the general relation

$$\langle \mathcal{F} e^{-\beta(\mathcal{W}-\Delta F)} \rangle = \langle \hat{\mathcal{F}} \rangle, \tag{6.7}$$

where $\langle \square \rangle$ signifies averaging over several trajectories, and $\mathcal{F}$ is some path function, e.g. heat or work.

This expression is referred to as *Crook's equation*, from which multiple important relations can be derived, such as the fluctuation theorems [103]. In particular, choosing $\mathcal{F} = \hat{\mathcal{F}} = 1$ we find

$$\langle e^{-\beta(\mathcal{W}-\Delta F)} \rangle = 1. \tag{6.8}$$

The free energy difference is path independent, so we can take it out of the average. Moving the free energy to the right hand side of the expression we find the *Jarzynski equality* [104]

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F} \tag{6.9}$$

This is a particularly useful result that has quickly gained a lot of traction within the statistical thermodynamics community [105]. It tells us that if we start out from an equilibrium state, and then many times consistently drive the system out of equilibrium according to some predefined protocol captured by $\lambda(t)$, while measuring the work along each corresponding trajectory, then we can estimate the free energy difference of that system. The free energy is generally a notoriously hard to quantify since it depends on the entropy, which is an elusive quantity. This is also a boon to experimentalists, since it is generally hard to perform experiments on a system while simultaneously keeping it in equilibrium. Using the Jarzynski equality we can then circumvent this issue altogether and retrieve equilibrium properties from out-of-equilibrium observations. This is exactly what we are going to do the the following sections.

6.2 The image reconstruction problem formulation

A challenging aspect of reconstructing surfaces in both physical and simulated AFM experiments is that the tip spends much more time near potential minima than maxima. A consequence of this is that we will obtain a disproportionately large amount of data about the sticking parts of the surface, and conversely very sparse data about the slipping parts. Since we will be lacking data density in these regions then, we can expect the reconstruction here to be less reliable. A way to rephrase this issue is that the temporally indexed data that we receive from experiments, badly corresponds to the spatially indexed data that we seek. Our task

then, is to try to work with, or work around, this limitation when trying to reconstruct an image of the surface.

Consider the following scenario: A particle is at equilibrium positioned in some minimum of a potential energy. Given sufficient amount of time, the particle will explore the entire phase space available to it due to thermal fluctuations. Hence, if we wait and monitor the correct observables of the particle during this process, we will, after some arbitrary amount of time, be able to reconstruct the potential energy, and therefore the underlying surface. The problem is of course that: "sufficient amount of time" can be long enough that "waiting" loses all its practical meaning.s This conundrum is quickly solved by forcing the particle out of its equilibrium using some biasing potential, this is referred to as umbrella sampling[1]. Notably, a simple way of achieving this is to couple the particle to some moving support, and so we have – in the case of a simple one dimensional corrugated potential – retrieved the PT-model.

If we perform multiple such experiments, and collect the data from these, we should be able to reconstruct the underlying potential energy using clever tricks from statistics and probability theory. We will now look into three such methods: The weighted histogram method, the Lucy-Richardson deconvolution, and explicit current methods. We will begin by separately formulating each of the methods, review how they operate, and benchmark their performance. Then we will compare and comment on the results.

6.3 The weighted histogram method

The first of the methods that we will consider is the Weighted Histogram Method (WHAM) [107]. Let $U(x,t)$ be a biasing potential as it was introduced above. We can discretize this to obtain a set of biasing potentials of the form $V_i(x) = V(x,t_i)$, that is, we treat the biasing potential obtained at every timestep as separate biasing potentials. Furthermore

[1]We would like to make a honourable mention of Naughton's *Coding and Cats* [106] at this point for impeccably illustrating umbrella sampling with cats and boxes.

let p_{ij} be the biased probability that x is binned into the jth bin under the biasing potential $V = V_i + U^0$, where we have introduced U^0 to designate the unbiased potential. This is the notation that we will use consistently for the WHAM indexes: i indexes the biasing potentials (in terms of the support position), whereas j indexes bins along the x-axis. We can relate this biased probability to the unbiased probability by

$$p_{ij} = f_i c_{ij} p_j^0, \tag{6.10}$$

where p_j^0 indicates the unbiased probability of finding the tip in bin j, c_{ij} is some biasing factor for the specific combination of bin and biasing potential, and f_i is a normalization factor such that if we sum over all bins

$$\sum_j p_{ij} = 1, \tag{6.11}$$

meaning,

$$f_i = \frac{1}{\sum_j c_{ij} p_j^0}. \tag{6.12}$$

If we only have data from one trajectory, then the best estimate of the unbiased potential in bin j given a biasing potential U_i can be calculated from the number of hits, n_{ij}, in that bin and the total number of datapoints N_i as

$$\overline{U}_{ij} = \frac{n_{ij}}{N_i f_i c_{ij}}, \tag{6.13}$$

where we will take $\overline{\square}$ to mean estimate.

The idea here is to calculate the probabilities for the tip to end up in some bin, and from this distribution infer the shape of the underlying potential energy, this is illustrated in figure 6.2. WHAM then, is the

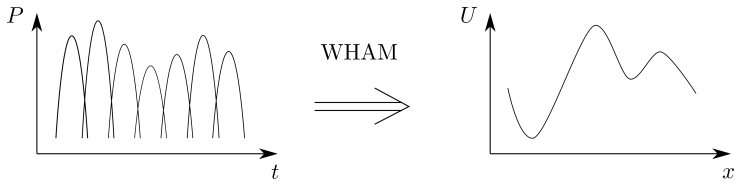


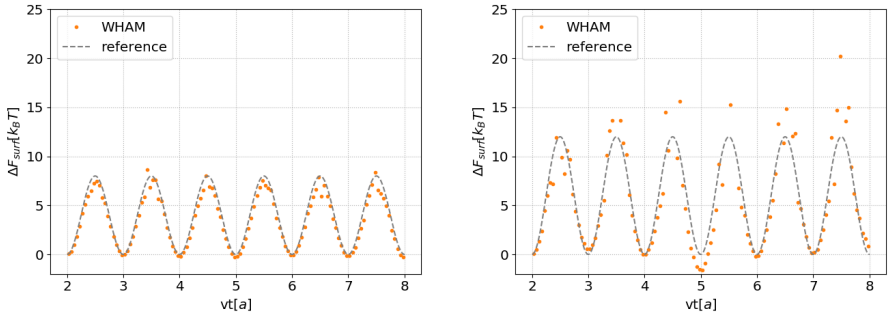
Figure 6.2: Illustrating the WHAM method, where the potential energy surface is found from binning the average work done on a system.

answer to how these weights c_{ij} should be calculated as well as how the data across multiple trajectories should be combined to give the best estimate for the unbiased potential [107]. For practical reasons, what we will bin is not the exact position of the tip, but rather the instantaneous work done by the system. It has been shown that the Jarzynski equality can be used in conjunction with WHAM to yield a free energy surface estimate as [108]

$$F_0(z) = -\beta^{-1} \log \frac{\sum_i \frac{\langle \delta(z - z_i) e^{-\beta w_i} \rangle}{\langle e^{-\beta w_i} \rangle}}{\sum_i \frac{e^{-\beta V_i(z,\lambda)}}{\langle e^{-\beta w_i} \rangle}}, \tag{6.14}$$

where z is a coordinate, λ is the protocol for driving the system out of equilibrium, $V(z, \lambda)$ is the biasing potential, w_i are weights, and δ is the Kronecker delta.

In our case, the coordinate is the position of the tip, x, the protocol is as mentioned earlier $\lambda = vt$, the biasing potential is $V = \frac{k}{2}(x - vt)^2$, and the weights are the sums of the work in each bin $W_j = \frac{\sum_i W_{ij}}{N}$, where N is the number of traces. Meaning our free energy estimate becomes



(a) WHAM for a barrier height of $8k_BT$

(b) WHAM for a barrier height of $12k_BT$.

Figure 6.3: Results from image reconstructions from AFM data using a periodic substrate while changing the barrier height. We observe how WHAM struggles for high energies.

$$F_0(x) = -\beta^{-1} \log \frac{\sum_i \frac{\langle \delta(x - x_i) e^{-\beta W_i} \rangle}{\langle e^{-\beta W_i} \rangle}}{\sum_i \frac{e^{-\beta k/2(x - vt_i)^2}}{\langle e^{-\beta W_i} \rangle}}. \tag{6.15}$$

This expression looks opaque, but it is really just the exponentiated work in a bin, averaged over all the trajectories, weighted by the scaled biasing potential, then we take the logarithm of this to get rid of the exponential function. Notice how WHAM becomes a tool to connect time distributed data (work at specific times), and data distributed over position (free energy).

We present data from WHAM reconstructions in figure 6.3, obtained from repeated *in silico* AFM experiments using the Langevin dynamics and the PT-model. The energy barrier height is given in k_BT, and the sliding distance is given in lattice parameters. Each reconstruction is the result of 30 traces. As can be seen WHAM performs reasonably well for low energies, but breaks down for high barrier heights.

We can understand the reason behind WHAM's shortcomings by recalling the definition of the WHAM free energy estimate (eq. 6.15). In the numerator, the binning of works appears as $\langle \delta(x - x_i) e^{-\beta W} \rangle$, an estimate of this work distribution after N experiments is

$$\langle \delta(x - x_i) e^{-\beta W_j} \rangle_j = \frac{1}{N} \sum_{i=1}^{N} K(x_i) e^{-\beta W_j}, \tag{6.16}$$

where j enumerates bins, i enumerates traces, and $K(x_i)$ is the indicator function

$$K(x_i) = \begin{cases} 1 & \text{if } x_i \in b_j, \\ 0 & \text{if } x_i \notin b_j \end{cases}. \tag{6.17}$$

where b_j is the bin labeled j.

This expression estimates the average exponentiated work in a bin by averaging over all exponentiated works across all experiments which were sorted into that particular bin. A dire consequence of this is that if there are no hits in a bin, that bin will not contribute to the overall work estimate at all. This is particularly pathological for stick-slip motion, since we rely on capturing rare high slipping states that contributes significantly to the average energy. Furthermore WHAM includes no way of mitigating this by e.g. interpolation to approximate these points. The result of this is the complete failure of WHAM to reconstruct the potential energy in bins with no hits, and a very poor performance in bins where the data is sparse enough. As if that was not bad enough in itself, our troubles with WHAM do not end there. The sole source of energy dissipation in the PT model comes from viscous damping. This damping is only tangible when the tip is moving fast – that is ,during slips. Meaning, WHAM will miss most of the energy dissipation in the system leading to artifacts in the form of an apparent cumulative buildup of excess energy with each slip, this can be seen in figure 6.3b. This renders WHAM inapt for reconstructing highly corrugated potential energy

surfaces using an AFM-like approach, and we shall have to move on to a more sophisticated method.

6.4 The Lucy-Richardson deconvolution

If we for a moment go back to the problem formulation as one where we have to reconstruct an image by extrapolation from incomplete data, then aid may come from an unexpected direction. As it turns out, this is not at all an unfamiliar situation to astronomers who are *constantly* challenged with extracting as much meaningful information as possible from limited information (light) sources[2].

Assume that we have some detector that is gathering incandescent photons on some grid of cells (pixels). Then there is always in practice a non-zero probability that, for one reason or another, a photon that was *supposed* to end up in some pixel x, instead ends up in some nearby pixel x'. This is what we would commonly refer to as noise. The resulting blurred image can be described by the so called imaging equation, which captures the convolution of the real image and the noise

$$I_j = \sum_i p_{ij} s_i, \tag{6.18}$$

where p_{ij} is the probability that a photon that should have been detected at pixel i is instead detected at pixel j due to noise, s_i is the real – or true – uncorrupted pixel data at pixel i, and I_j is the corrupted pixel data at pixel j – or, the observed image, as it were. In the terminology of image reconstructions p_{ij} is called a *point spread function* but is in some contexts referred to just as *the kernel*. It is generally taken to be Poisson distributed.

[2]These techniques are also widely known to radiologists working in medical imaging, but historically they too have learned many of their tricks from the astronomers [109], so credit where credit is due.

The "image" we can readily obtain from AFM experiments is the work done by the support as the tip is being pulled. This then – the averaged work after many trajectories – will be our input-signal. We can write this in terms of the Jarzynski equality (eq. 6.9) as

$$I_j = \langle e^{-\beta W} \rangle_j, \tag{6.19}$$

here we implicitly discretized the Jarzynski equality such that the pixel j in the imaging equation now corresponds to the average exponentiated work in bin j, using a similar binning approach as with umbrella sampling as for WHAM.

Furthermore, we know that the real image in the AFM case is the surface potential, and also that the surface potential is distorted by the moving support, meaning, we have a (discretized) convolution like

$$\langle e^{-\beta W} \rangle_j = \sum_i e^{-\beta V(i,j)} e^{-\beta U_s(i)}, \tag{6.20}$$

where U_s is the surface potential, and V is the biasing potential, here i indexes the position of the tip.

But, we already established that the left hand side of this equation is the image in the imaging equation, and moreover, the right hand side is a convolution of the biasing and surface potentials. Meaning, we can identify these as the point spread function and surface potential respectively

$$p_{ij} = e^{-\beta V(i,j)} = e^{-\frac{k}{2}(x_i - v t_j)^2}, \tag{6.21}$$

$$s_i = e^{-\beta U_s(i)}, \tag{6.22}$$

and we have successfully related the imaging equation to the AFM system. Collecting this into one equation, we have

$$I_j = \sum_i p_{ij} s_i \iff \langle e^{-\beta W(x)} \rangle_j = \sum_i e^{-\frac{k}{2}(x_i - vt_j)^2} e^{-\beta U_s(i)} \tag{6.23}$$

where i enumerates trajectories, and j indexes the bins.

Our task is to find a way to solve this expression for $U_s(x)$. However, as shown in references [110, 111], solving such an equation in the presence of stochastic noise is an ill-posed problem[3]. Hence, we will have to resort to some numerical algorithm for finding an approximate solution. The Lucy-Richardson Deconvolution (LRD) is just such an algorithm [112][113]. A contemporary report on LRD is available here [114] for reference. LDR is an iterative method which in a discretized form reads [115]

$$s_i^{k+1} = s_i^k \sum_j \frac{I_i p_{ij}}{\sum_i p_{ij} s_i^k}, \tag{6.24}$$

where s_i^k is the obtained estimate of the real image after k iterations in bin i and I_i is the initial image in bin j.

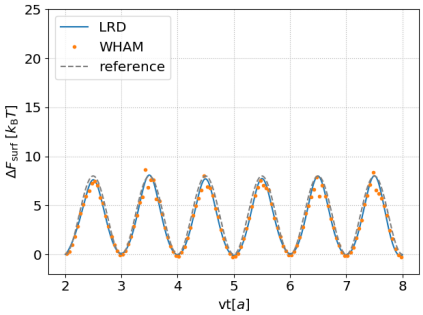
We can rewrite this expression in perhaps a slightly more intuitive way in vectorized form

$$\mathbf{S}^{k+1} = \mathbf{S}^k \left(P^{\mathrm{T}} \cdot \left(\frac{\mathbf{I}}{P \cdot \mathbf{S}^k} \right) \right), \tag{6.25}$$

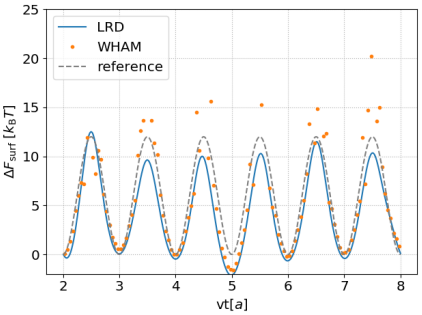
where division implies elementwise division and T denotes transpose, and I, P and S like before are the initial image, point spread function and the current image estimate respectively.

We present data from LRD reconstructions in figure 6.4, and compare this to the reconstruction performance of the WHAM algorithm. The

[3] A problem is ill-posed if it does *not* satisfy at least one of the following: 1. The problem has a solution, 2. The solution is unique, or 3. The solution depends continuously on the data.



(a) LRD and WHAM for a barrier height of $8k_BT$



(b) LRD and WHAM for a barrier height of $12k_BT$.

Figure 6.4: Results from image reconstructions of AFM data using a periodic substrate while changing the barrier height. We observe how LRD out performs WHAM, however, LRD is also starting to exhibit numerical instabilities in the high energy example.

reconstruction data here is the same as in figure 6.4. We see how LRD fares comparatively well to WHAM even high higher energies, however, it is also evident that LRD is on the verge of failing.

A well-known issue with LRD is that it deals poorly with low signal-to-noise ratios [116], in these cases it can result in what is known as noise amplification or over-fitting, where the algorithm will in fact start treating noise as signal components and thus counter-productively start blurring signal. There are various ways of mitigating this – the most efficient one (if circumstances permit) of course being obtaining more data less polluted by noise, but barring that, we have used stopping criteria (we simply used 30 LRD iterations) to interrupt the algorithm before the noise became too pronounced [116, 115].

A bit atypically, the reason for this is however not related to the noise. In fact, the thermal noise itself has constant amplitude since we keep the temperature fixed, and moreover, since we are increasing the barrier height, this ratio actually decreases. Rather, our issue is once again sparse data. LRD applies its deconvolution filter indiscriminately across

all points, even the ones where no data was obtained. In some sense, this is not a bug, it is a feature. LRD has some manner of interpolation built-in interpolation (c.f. the matrix operations in 6.25), and this is the main reason it out-performs WHAM, however, for sparse enough data, this auxiliary interpolation breaks down, and is no longer able to salvage the missing data.

Thwarted once again by the disproportionate availability of data, we shall abandon trying to apply already established general image reconstruction protocols to AFM data, and instead make an effort to design a custom made algorithm, specialized for the situation we are investigating.

6.5 Explicit current estimation methods

In WHAM we investigated how the distribution of work in the system could be used to infer information about the underlying potential energy. In a sense, this is only one example of a larger class of image reconstruction methods looking at the flow of dynamic quantities in the system through binning across multiple repeated experiments. At present time, this is where we are focusing our efforts. We have some sophisticated ideas about how to accomplish this, but since this is a work in progress, we will only review the most elementary case here as a proof of concept, and comment on what ways we are looking into at the moment to expand upon this.

The basic idea here is to look at the Fokker-Planck equation. This is a partial differential equation that describes how a probability distribution of some system evolves with time under some stochastic perturbations to its dynamics (see [117] for a formal treatment, and [118] for an undergraduate level treatment). Because of this quite general definition, the actual Fokker-Planck equation in one system might not look much like the Fokker-Planck equation in another system.

The equations of motion for an overdamped PT system is

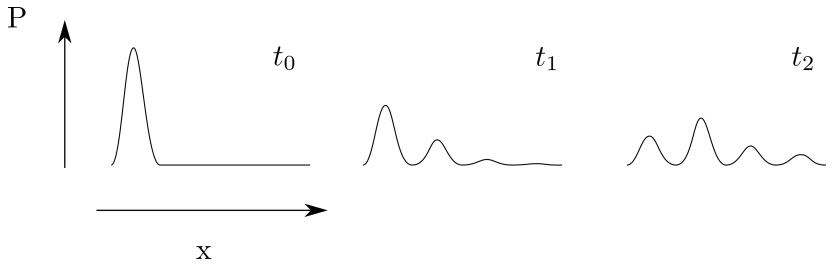


Figure 6.5: An illustration of the time evolution of the probability distribution corresponding to the PT-model. The distribution captures the presumptive position of the tip as a function of time.

$$\dot{x}(t) = -\frac{1}{\eta}\nabla_x U(x) - \frac{k}{\eta}(x(t) - vt) + \sqrt{\frac{2k_B T}{\eta}}\,\xi(t), \tag{6.26}$$

where the notation is identical to the other instances of Langevin dynamics we have studied in this thesis.

The corresponding Fokker-Planck equation becomes

$$\frac{\partial p}{\partial t} = -\nabla_x \left(\frac{1}{\eta}\nabla_x U(x) - \frac{k}{\eta}(x(t) - vt) - \frac{k_B T}{\eta}\nabla_x \right) p(x,t), \tag{6.27}$$

here p(x,t) is the probability distribution telling us what the probability is of finding the tip in the position x at the time t.

It is common – and useful – to regard the right left side of the Fokker-Planck equation as an evolving *probability current*

$$\frac{\partial p}{\partial t} = -\nabla_x \boldsymbol{j}(x,t). \tag{6.28}$$

This gradient of the current then describes how the probability of finding the tip at a specific position at a specific time changes with time.

We supply a figure (fig. 6.5) to give a schematic overview of how this probability might evolve in the PT-model. In this figure, we let a PT-model start in equilibrium, meaning that the tip is safely tethered to the first minimum. Then we start moving the support. This will mean that we – across many simulations – no longer know where we will find the tip. However, we know that the probability of finding the tip in the first minimum will decrease, furthermore we know that the probability of finding it in the second minimum consequently will increase, since there is a non zero probability that it will have slipped over into this minimum. At every time, we know where the support is, since it moves with constant velocity, but as time flows, the probability distribution describing where to find the tip will become increasingly complex. The Fokker-Planck equation describes how this probability distribution changes with time. And the probability current gives the instantaneous change of the probability distribution at some time and position.

This then challenges us with the question: If we bin the positions of the tip at some time intervals across multiple simulations to retrieve the probability distribution of the tip's position over time, is it possible to reconstruct the underlying potential energy that gave rise to the particular distribution found? This is indeed the case, and we shall here review the naive way of doing so.

Assume that the estimated discretized (henceforth denoted $\hat{\square}$) probability distribution can be described by some kernel function K

$$\hat{p} = \frac{1}{N} \sum_{i=0}^{N} K(x_t^{(i)}, x), \tag{6.29}$$

here K takes the arguments x being some tip position, and $x_t^{(i)}$ being the position of the tip in the trace indexed by i at time t. For a simple uniform binning, K can be taken to be the indicator function

$$K(x_t^{(i)}, x) = \begin{cases} 1 & \text{if } \{x_t^{(i)}, x\} \in b_k, \\ 0 & \text{if } \{x_t^{(i)}, x\} \notin b_k, \end{cases} \tag{6.30}$$

where b_k is a bin. The indicator function then simply returns 1 if x shares bin with $x_t^{(i)}$, and the interpretation of 6.29 becomes the fraction of times the tip was in bin k_k.

Using this estimate of the probability distribution we can estimate the corresponding probability current by simply counting the average number of transitions over some interval $[x, x + \Delta x]$ likeso

$$\hat{j}(x,t) = \frac{\hat{p}(x,t)}{\Delta t} \frac{\sum_{i=1}^{N-1} K\left(x_t^{(i)}, x\right)\left(x_{t+\Delta t}^{(i)} - x_t^{(i)}\right)}{\frac{1}{N}\sum_{i=1}^{N-1} K\left(x_t^{(i)}, x\right)}. \tag{6.31}$$

Now we can go back to our original Fokker-Planck equation (eq. 6.27), and combine it with equation 6.31 to obtain the following expression

$$\nabla_x \hat{U}(x,t) = \frac{\hat{j}(x,t) + \frac{k_B T}{\eta}\nabla_x \hat{p}(x,t)}{\hat{p}(x,t)} - \frac{k}{\eta}(x - vt), \tag{6.32}$$

for the force profile of the potential energy.

The potential energy can then be obtained by integrating this expression with respect to x, which is easily done by conventional numerical integration techniques.

At present, we have not yet implemented this method for any physical data. However, it has shown promise in trial runs on reduced artificial datasets. Since these results are preliminary, and that they are not comparable to the LRD and WHAM results at this stage, we have omitted including them in this thesis.

6.6 Discussion and outlook

Summarizing, WHAM, LRD, and the explicit current estimation method all aim at solving the same problem – reconstructing an image. However, the approaches of the methods are essentially different: WHAM relies

on knowing how the system was driven out of equilibrium, and then trying to undo the effects of this; current estimation have similarities with WHAM in that it is a binning method estimating a probability distribution, but it requires no biasing, rather it works directly on the data; LRD takes a different approach in that it is an iterative method relying on global transformations of the data rather than local isolated reconstructions.

The issues of WHAM and LRD are hard to engineer away. Since both are failing due to lack of data and not due to e.g. high noise intensity, and furthermore – given the improbability of the missing data – that obtaining more data is computationally expensive and perhaps even experimentally impossible, we are not very optimistic about the prospects of using these methods for reconstructions in AFM experiments.

We are however optimistic about reconstructing free energy surfaces using probability currents. On the onset, this approach suffers the same shortcomings as previous models, estimating the current accurately requires vast amounts of data to compensate for improbable events. However, recent developments in [119] show that it is possible to use deep learning to train a neural network to optimize a model for entropy production in an AFM system. Incidentally, the cost function in this optimization protocol is proportional to the probability current of the position. This reduces the data required since the neural network makes more efficient use of the available data through interpolations and sophisticated fitting of local parameters to react to changes in the data. We are collaborating right now to achieve this.

Part IV

Epilogue

Concluding reflections

This is the part where one would typically tie up the red thread coursing through this thesis in a tidy knot and show how all the results fit together to form a single unifying picture. At first glance, tying together the nanofriction of atomically thin sheets, the social dynamics of common-pool resource games, and the ranking of image reconstruction methods, surely would take politician levels of word-gymnastics. This is not an entirely fair representation however. Granted, while the fields spanned by this thesis indeed are wide, the tools with which we analyzed them were the same: simple models for out of equilibrium systems. Here then, I shall highlight some of the insights assembled in this thesis, and supply my personal reflections on their significance.

I started out this thesis analyzing what pieces were missing in the puzzle of friction modeling of sheets on the atomic scale. For a long time the PT-model had been widely implemented to describe friction on the atomic scale, why then was it unsuccessful in describing layered 2D materials in particular? People had many ideas about why: because the sheet is bending, because the corrugation changes, because of how individual atoms stick to the actual surface of the AFM tip. Fundamentally, my contribution to this conundrum was entering with an open mind. I did not take sides in which detailed explanation was the correct one, rather I tried to capture the present dynamics from a high level perspective. I made some, in hindsight, pretty naive extensions to the PT-model, and then modified those based on the experimental context. From this I provided a coherent framework within which the established ideas about the origins of deviations from the PT model fit well. It is hard for me to tell exactly why no one had tried this quite simple approach before, but it illustrates an important point which is that in order to understand a system, sometimes it is helpful not to run all the way down the rabbit hole of one single aspect of the system.

Then came the large scale common-pool resource games, and the social dynamics of these. The story here goes much the same: it started out with a simple idea. The flow of opinions within a population seems to

be an intimidatingly hard thing to simulate, if nothing else simply due to the sheer number of sources of – and channels to – influence there are. There are incredibly complex models available for individual components of this influence, such as mass media. Having no background in neither social dynamics nor sociophysics, I definitely came into this field as an outsider, which meant that it took quite sometime to orient myself. Crucially however, I was not lost to blindly wander the lands of complexity science in solitude, I was guided by an expert in the field helping me work out which avenues probably would and would not be worth investigating. Still quite oblivious to what paradigms was dominating the field at the moment, I wrote down a simple model for how the dynamics might be governed, condensing as many complex factors as I could since I did not know how they were usually modeled on a detailed level, and then put my model to work. The result was a novel approach to model the rate at which consensus is reached in important society spanning coordination problems.

While the results presented in Part I and II definitely carries their own scientific merit, perhaps the most important message of this thesis lies in the intersection of these two projects on an abstract level and is really in terms of meta-science: *Sometimes when we try to understand a complex system we get lost in the anatomy of the problem – we want to understand the parts in order to build the whole as it were, but much like a fractal, sometimes describing one part reveals even more underlying complexity. In order to obtain a baseline of understanding, it can then be useful to adopt a qualitative approach and develop a simple model for the system.* I am not trying to overthrow reductionism here by ushering in an age of holism, but I would like to make a case for the usefulness of simple models as a tool that might otherwise become lost to time in the era of detailed computer simulations and narrow research questions.

Obtaining accurate qualitative information is challenging, because "qualitative" means stepping away from the details, whereas "accurate" typically involves getting more into the details. It is therefore important that qualitative ideas originate in the intersection of high level understandings of the different aspects of a system. Based on this I would also

like to promote multidisciplinary collaborations in science, as a means to further develop the qualitative understanding of complex systems.

Finally, in Part III, I abandoned the bird's-eye view to actually get nitty-gritty in the details about image reconstruction algorithms. While this project did not bear fruit in time to largely contribute to the scientific value of this thesis, a significant amount of time (paradoxically mostly at the early stages of the PhD) was spent here and I felt compelled to report it as-is. This project did not quite turn out the way I thought it would, and now I am investigating developing a new image reconstruction technique rather than using an established one. I am optimistic about the prospects of this, but time will have to be the judge of its ultimate success.

To any reader that made it this far I would like to extend my heartfelt gratitude for taking the time to read this thesis. I hope it was an enjoyable – or at least illuminating – read, bringing you some new insights about the topics discussed within. I will now conclude this thesis with some acknowledgments.

Thank you

In this section I wish to acknowledge some of the remarkable people and institutions which I have had the great privilege and honour of knowing, working with, and learning from during the course of my time as a PhD student. Each person will be addressed in a highly personal manner, and reading these statements might not make much sense to a reader who either are not one of them or at least know one of them. Regardless, they are all meant for their respective recipient, so that is all in order. With that said, any reader who disapproves of emotional sentiments should count themselves warned!

I have always surrounded myself with plenty of people, and the road to this PhD has been a lifelong journey, so there are many people to address here.

To my parents, I am eternally grateful for all the love and support that you have given me through all the years I have pursued this goal. The gods know it has not been a straight road, but in hardship and in celebration, you have encouraged me all the way. During this journey you have truly given meaning to the phrase "unconditional love".

To my brother, the funny guy, the prankster, the stoic. You have always been a walking paradox of decorum and absolute hedonism. Thank you for providing me plenty of both during my time as a PhD. Moreover thank you for the brotherhood that we share, maybe no one else will ever understand it, but they don't have to!

Supriya, to my co-supervisor I wish to say this: I'm regretful that we did not get to spend more time together than we did. Every time I met with you in Stockholm I got the distinct feeling that I had a lot to learn from your tutelage. I hope we shall find time to make up for this in the future. Specifically I want to thank you for moral support during the final stages of my Phd.

Robin, my partner in crime, "PhD sibling", and steadfast travel companion. Many adventures did we experience during our time together in Trondheim. And in Trieste. And in Osaka. And in... etc. You always had my back, and I always had your. Much have we shared, and much shall we still share together. Thank you for being the friend I needed in a town where I knew no one.

Jenny, you know how you sometimes meet someone, and you just know: "yeah, that one right there – that's one of the good 'uns". Well, that's the distinct feeling I got when you strode into the department with those sassy steps of yours. No matter if it's gifs, gossip, or goofing, you're the girl! I can't imaging ever having a dull moment with you around. Thank you for being you, you are truly an amazing piece of mankind! On a more professional note, thank you for our collaboration on the "clept" paper, I have thoroughly enjoyed our time together by the white board.

Anders, this, our graduation as PhDs, is our joint victory to celebrate together. For the last 10 years you and I have been virtually inseparable. Exams, lab work, projects, presentations, course work, theses, we have faced it all together, and we have prevailed. I owe so much to you, being here today. You have my unending gratitude, and my unwavering friendship. Thank you.

Jesper, flat-mate, code mate, climb mate, great mate. Thank you for all the good times we have shared, and for throwing a rope down some of the deeper coding pits. You are a loyal friend, and I shall never forget all the time we spent together.

Signe, we have shared some of the best and worst moments of my still short life together. Thank you for taking a leap of faith with me to support my dreams. Furthermore, thank you for many enlightening and enlightened discussions about just about everything, in particular per-

haps all the ideas that led to the collective action paper.

Robert and Yi-Hua, long have we ventured since we extracted kiwi DNA and made thermite in that basement lab where we all met – and yet, still long is our journey ahead of us. You, my faithful companions, have been with me since my very first steps on this PhD journey. Thank you for company, inspiration, mad projects, and your support. Together we shall continue to leave our marks in the annals of history, one achievement at the time.

Sreekanth, thank you for our work together on image reconstructions and for being very considerate during my end-of-PhD stress. I'm looking forward to working more with you in the future. Moreover, thank you for our brief work together on FysikShow, we were – of course – spectacular!

Sigrid, the perceptive master student. Thank you for our collaboration on the collective action paper. Much did we learn together, and great things came out of it. Being your co-supervisor was my great privilege.

Drew – the cool guy – if I ever had a science idol, you probably came the closest! Worked on hottest topics, perpetually sarcastic, and with great personal charisma to boot. Collaborating with you on the collective action paper has been an eye opener that really inspires me to continue pursuing a career in science.

Alexandra, who did not shy away from the challenge when I asked her to contribute cover art representing AFMs, social network dynamics and image reconstruction methods. You have, as anyone can see on the cover of this thesis, made a marvelous job, exceeding my already high expectations.

Ewa and Mia, the enablers. Being a kid struggling in school with dreams of understanding how the world worked, I always seemed to have the drive, but never the tools, to become a scientist. Ewa and Mia provided me with the environment I needed to make those tools, and they taught me how to use them. We must never forget those who inspired us to become what we are – we must never forget our teachers. Thank you so much!

The tribo community, during every single conference, workshop, and summer school I have participated in during my short but intense scientific career, they have all had one thing in common: an openness, a curiosity, a welcoming atmosphere. No matter if it has been an interaction with a fellow PhD student like Zazo Meijs, or a "big shot" such as Robert Carpick, I have consistently felt like I have been greeted as an equal – *as another guy in the tribo-gang.* I have shared food, animated scientific discussions, and yes, elephant tours with people from all levels, all nationalities, and all backgrounds – and it has been the most natural thing. To the international tribology community, I take off my hat, and I salute you all. The world of science would be a better place if all communities were as welcoming as you are.

The COST action MP1303 was an initiative organized by European tribologists to create a platform for exchanging ideas about, arranging training in, and advancing the field of, tribology. It was funded by the EU, and helped hundreds of young tribologists like myself to establish themselves as internationally competitive contributors to the field. Specifically, MP1303 organized and paid for my participation in a tribology summer school at DTU in Copenhagen 2017. MP1303 also helped me promote my research at my first scientific conference: *Trends in Nanotribology* 2017 in Trieste, from where I have many cherished memories. I wish to extend my heartfelt appreciation to MP1303 and the organizers of these events for supporting me and advancing my, and my fellow young researchers', scientific careers.

Gurli Grundströms Forskarstipendiefond, is a memorial foundation that generously supports young researchers by providing travel grants. This foundation helped me realize a research visit to Tel Aviv University in 2018 where me and my supervisor laid the groundwork for a new research project with our collaborators there. It was an inspiring visit that furthered our collaboration, and I am thankful for Gurli Grundströms Forskarstipendiefond having provided me with this support. Moreover, I am grateful to **Michael Urbakh** for receiving me and, in true international tribology community manner, making me feel right at home

Coding support volunteers, I had barely written a line of proper computer code in my life when I started this PhD. Yet, here I am, tens of thousands of lines of code later. I owe much of my quick progression into a confident programmer to the thousands of people out there that graciously volunteer their time and resources into supporting more junior programmers through initiatives such as StackExchange, Geeks for Geeks, etc. I have every intention of returning the favour by now in turn helping those in similar situations as I was four years ago.

The FOSS community, it is quite astonishing how much the science community at large owes to the free open source software community at this point. So intertwined are these communities today that it is hard to tell where one ends and the other begins. I'm no exception here, the FOSS community has provided me, free of charge, literally all the software I have crucially relied on during my PhD. I want to thank the many enthusiasts and supporters of this community for their contribution to this, and other research projects, for they hardly get the recognition they deserve.

Unga Forskare, the organization that fanned the flames of my scientific interest into a raging inferno. To be given a platform in which not only to exercise an interest in science at a young age, but also meeting other outliers like me wanting to do so is nothing short of priceless. Thank

you everyone that shared this experience with me, you are too many to list here. Keep on fighting for making science as natural a hobby for kids as football. Great things can come out of this – I should know!

Vetenskapsrådet, second to last but not least, I want to extend my warmest gratitude to Vetenskapsrådet for financing my PhD position through the grant 2015-04962, it is a dream come true and a life long goal achieved to get the opportunity to do a PhD. I am humble and grateful for having been given this opportunity.

Astrid, last but oh so very much not least, we arrive at you Astrid – where it started and where it ends. There really is a million things to say, but I shall have to keep it brief as I can. We make a pretty great team you and I. I think I owe a lot of the success I have had as a PhD student to that simple fact. I believe that a lot of supervisors would have "shot down" a lot of the ideas that I have had during the course of my PhD in order to "keep me on track". Quite the contrary, you have complimented my speculative ideas, carved away all the crazy stuff until there's a solid idea to be found, and then you have shot back with your own out of the box ideas – and it has served us well, I think this has laid the groundwork for a lot of the brave research projects we have pursued together. Thank you for providing a flexible environment where I have been able to use *all of my strengths* and grow as both a person and a scientist. Beyond having been an ideal supervisor for me personally, you are a role model for leadership in science. I dearly hope that more people shall take after your rare combination of scientific strategizing and caring deeply about the people around you. I know that I speak for both myself and my fellow PhD students when I thank you for going over and beyond in your unyielding conviction to preserve and improve the well-being of the people you supervise.

Bibliography

[1] D. Andersson and A. S. de Wijn, "Understanding the friction of atomically thin layered materials," *Nature communications*, vol. 11, no. 1, pp. 1–7, 2020.

[2] J. Roadnight Sheehan, D. Andersson, and A. S. de Wijn, "Thermal effects and spontaneous frictional relaxation in atomically thin layered materials," *arXiv preprint arXiv:2012.00371*, 2020.

[3] D. Andersson, S. Bratsberg, A. K. Ringsmuth, and A. S. de Wijn, "Dynamics of collective action to conserve a large common-pool resource," *arXiv preprint arXiv:2012.00892*, 2020.

[4] R. Munroe, "xkcd (no. 669) experiment," 2009.

[5] K. Holmberg and A. Erdemir, "Influence of tribology on global energy consumption, costs and emissions," *Friction*, vol. 5, pp. 263–284, Sep 2017.

[6] G. B. D. of Education and Science, *Lubrication (tribology) education and research : a report on the present position and industry's needs / Deartment of education and science.* H.M.S.O London, 1966.

[7] J. Williams, *Engineering tribology.* Cambridge University Press, 2005.

[8] M. Feldmann, D. Dietzel, H. Fuchs, and A. Schirmeisen, "Influence of contact aging on nanoparticle friction kinetics," *Physical review letters*, vol. 112, no. 15, p. 155503, 2014.

[9] Q. Li, T. E. Tullis, D. Goldsby, and R. W. Carpick, "Frictional ageing from interfacial bonding and the origins of rate and state friction," *Nature*, vol. 480, no. 7376, pp. 233–236, 2011.

[10] L. Prandtl, "Ein gedankenmodell zur kinetischen theorie der festen körper," *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 8, no. 2, pp. 85–106, 1928.

[11] G. T. B.Sc., "Cvi. a molecular theory of friction," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 7, no. 46, pp. 905–939, 1929.

[12] V. L. Popov and J. Gray, "Prandtl-tomlinson model: History and applications in friction, plasticity, and nanotechnologies," *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 92, no. 9, pp. 683–708, 2012.

[13] A. Vanossi, N. Manini, M. Urbakh, S. Zapperi, and E. Tosatti, "Colloquium: Modeling friction: From nanoscale to mesoscale," *Rev. Mod. Phys.*, vol. 85, pp. 529–552, Apr 2013.

[14] S. Y. Krylov, K. B. Jinesh, H. Valk, M. Dienwiebel, and J. W. M. Frenken, "Thermally induced suppression of friction at the atomic scale," *Phys. Rev. E*, vol. 71, p. 065101, Jun 2005.

[15] E. Gnecco, R. Bennewitz, T. Gyalog, C. Loppacher, M. Bammerlin, E. Meyer, and H.-J. Güntherodt, "Velocity dependence of atomic friction," *Phys. Rev. Lett.*, vol. 84, pp. 1172–1175, Feb 2000.

[16] K. B. Jinesh, S. Y. Krylov, H. Valk, M. Dienwiebel, and J. W. M. Frenken, "Thermolubricity in atomic-scale friction," *Phys. Rev. B*, vol. 78, p. 155440, Oct 2008.

[17] Y. Sang, M. Dubé, and M. Grant, "Thermal effects on atomic friction," *Phys. Rev. Lett.*, vol. 87, p. 174301, Oct 2001.

[18] A. Vakis, V. Yastrebov, J. Scheibert, L. Nicola, D. Dini, C. Minfray, A. Almqvist, M. Paggi, S. Lee, G. Limbert, *et al.*, "Modeling and simulation in tribology across scales: An overview," *Tribology International*, vol. 125, pp. 169–199, 2018.

[19] I. Szlufarska, M. Chandross, and R. W. Carpick, "Recent advances in single-asperity nanotribology," *Journal of Physics D: Applied Physics*, vol. 41, no. 12, p. 123001, 2008.

[20] D. Frenkel and B. Smit, *Understanding Molecular Simulation*. Orlando, FL, USA: Academic Press, Inc., 2nd ed., 2001.

[21] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *J. Comput. Phys.*, vol. 117, pp. 1–19, Mar. 1995.

[22] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, pp. 19–25, Sept. 2015.

[23] C. Lee, Q. Li, W. Kalb, X.-Z. Liu, H. Berger, R. W. Carpick, and J. Hone, "Frictional characteristics of atomically thin sheets," *Science*, vol. 328, no. 5974, pp. 76–80, 2010.

[24] T. Filleter, J. L. McChesney, A. Bostwick, E. Rotenberg, K. V. Emtsev, T. Seyller, K. Horn, and R. Bennewitz, "Friction and dissipation in epitaxial graphene films," *Phys. Rev. Lett.*, vol. 102, p. 086102, Feb 2009.

[25] S. Li, Q. Li, R. W. Carpick, P. Gumbsch, X. Z. Liu, X. Ding, J. Sun, and J. Li, "The evolving quality of frictional contact with graphene," *Nature*, vol. 539, pp. 541 EP –, Nov 2016.

[26] Q. Li, C. Lee, R. W. Carpick, and J. Hone, "Substrate effect on thickness-dependent friction on graphene," *physica status solidi (b)*, vol. 247, no. 11-12, pp. 2909–2914, 2010.

[27] Z. Ye, C. Tang, Y. Dong, and A. Martini, "Role of wrinkle height in friction variation with number of graphene layers," *Journal of Applied Physics*, vol. 112, no. 11, p. 116102, 2012.

[28] S. Kwon, J.-H. Ko, K.-J. Jeon, Y.-H. Kim, and J. Y. Park, "Enhanced nanoscale friction on fluorinated graphene," *Nano Letters*, vol. 12, no. 12, pp. 6043–6048, 2012. PMID: 22720882.

[29] X. Zeng, Y. Peng, L. Liu, H. Lang, and X. Cao, "Dependence of the friction strengthening of graphene on velocity," *Nanoscale*, vol. 10, pp. 1855–1864, 2018.

[30] A. S. de Wijn, "Flexible graphene strengthens friction," *Nature*, vol. 539, pp. 502 EP –, Nov 2016.

[31] M. Naylor, "Golden, and π flowers: A spiral story," *Mathematics Magazine*, vol. 75, no. 3, pp. 163–172, 2002.

[32] M. M. van Wijk, A. Schuring, M. I. Katsnelson, and A. Fasolino, "Relaxation of moiré patterns for slightly misaligned identical lattices: graphene on graphite," *2D Materials*, vol. 2, p. 034010, jul 2015.

[33] R. Shi, L. Gao, H. Lu, Q. Li, T.-B. Ma, H. Guo, S. Du, X.-Q. Feng, S. Zhang, Y. Liu, P. Cheng, Y.-Z. Hu, H.-J. Gao, and J. Luo, "Moiré superlattice-level stick-slip instability originated from geometrically corrugated graphene on a strongly interacting substrate," *2D Materials*, vol. 4, p. 025079, apr 2017.

[34] Z. Ye and A. Martini, "Atomistic simulation of the load dependence of nanoscale friction on suspended and supported graphene," *Langmuir*, vol. 30, no. 49, pp. 14707–14711, 2014. PMID: 25419859.

[35] Y. Dong, "Effects of substrate roughness and electron–phonon coupling on thickness-dependent friction of graphene," *Journal of Physics D: Applied Physics*, vol. 47, p. 055305, jan 2014.

[36] X. Zheng, L. Gao, Q. Yao, Q. Li, M. Zhang, X. Xie, S. Qiao, G. Wang, T. Ma, Z. Di, J. Luo, and X. Wang, "Robust ultra-

low-friction state of graphene via moiré superlattice confinement," *Nature Communications*, vol. 7, p. 13204, 2016.

[37] E. Pop, V. Varshney, and A. K. Roy, "Thermal properties of graphene: Fundamentals and applications," *MRS bulletin*, vol. 37, no. 12, pp. 1273–1281, 2012.

[38] K. Zakharchenko, M. Katsnelson, and A. Fasolino, "Finite temperature lattice properties of graphene beyond the quasiharmonic approximation," *Physical review letters*, vol. 102, no. 4, p. 046808, 2009.

[39] H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," *Physica*, vol. 7, no. 4, pp. 284–304, 1940.

[40] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: fifty years after kramers," *Reviews of modern physics*, vol. 62, no. 2, p. 251, 1990.

[41] J. Kurkijärvi, "Intrinsic fluctuations in a superconducting ring closed with a josephson junction," *Phys. Rev. B*, vol. 6, pp. 832–835, Aug 1972.

[42] L. Sjögren, "Lecture notes on statistical physics: Chapter 8," 2014.

[43] J. S. Langer, "Statistical theory of the decay of metastable states," *Annals of Physics*, vol. 54, no. 2, pp. 258–275, 1969.

[44] S. Hameroff and R. Penrose, "Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness," *Mathematics and computers in simulation*, vol. 40, no. 3-4, pp. 453–480, 1996.

[45] L. K. McKemmish, J. R. Reimers, R. H. McKenzie, A. E. Mark, and N. S. Hush, "Penrose-hameroff orchestrated objective-reduction proposal for human consciousness is not biologically feasible," *Physical Review E*, vol. 80, no. 2, p. 021912, 2009.

[46] G. Bianconi and A.-L. Barabási, "Bose-einstein condensation in complex networks," *Physical review letters*, vol. 86, no. 24, p. 5632, 2001.

[47] R. N. Mantegna and H. E. Stanley, *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.

[48] S. Kuhn, "Prisoner's dilemma," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Metaphysics Research Lab, Stanford University, winter 2019 ed., 2019.

[49] W. F. Lloyd, *Two lectures on the checks to population*. 1833.

[50] C. A. Holt and A. E. Roth, "The nash equilibrium: A perspective," *Proceedings of the National Academy of Sciences*, vol. 101, no. 12, pp. 3999–4002, 2004.

[51] G. Hardin, "The tragedy of the commons," *science*, vol. 162, no. 3859, pp. 1243–1248, 1968.

[52] J. B. Wiesner and H. F. York, "National security and the nuclear-test ban," *Scientific American*, vol. 211, no. 4, pp. 27–35, 1964.

[53] M. A. Nowak, "Five rules for the evolution of cooperation," *science*, vol. 314, no. 5805, pp. 1560–1563, 2006.

[54] N. Bostrom, "The vulnerable world hypothesis," *Global Policy*, vol. 10, no. 4, pp. 455–476, 2019.

[55] P. Friedlingstein, M. Jones, M. O'sullivan, R. Andrew, J. Hauck, G. Peters, W. Peters, J. Pongratz, S. Sitch, C. Le Quéré, *et al.*, "Global carbon budget 2019," *Earth System Science Data*, vol. 11, no. 4, pp. 1783–1838, 2019.

[56] M. Scheffer, S. R. Carpenter, T. M. Lenton, J. Bascompte, W. Brock, V. Dakos, J. Van de Koppel, I. A. Van de Leemput, S. A. Levin, E. H. Van Nes, *et al.*, "Anticipating critical transitions," *science*, vol. 338, no. 6105, pp. 344–348, 2012.

[57] S. M. Gardiner, *A perfect moral storm: The ethical tragedy of climate change*. Oxford University Press, 2011.

[58] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Advances in Complex Systems*, vol. 3, no. 01n04, pp. 87–98, 2000.

[59] R. Hegselmann, U. Krause, *et al.*, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of artificial societies and social simulation*, vol. 5, no. 3, 2002.

[60] T. Li and H. Zhu, "Effect of the media on the opinion dynamics in online social networks," *Physica A: Statistical Mechanics and its Applications*, p. 124117, 2020.

[61] J. Lorenz, "Continuous opinion dynamics under bounded confidence: A survey," *International Journal of Modern Physics C*, vol. 18, no. 12, pp. 1819–1838, 2007.

[62] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the national academy of sciences*, vol. 99, no. suppl 3, pp. 7280–7287, 2002.

[63] D. Agrawal and L. Bhuyan, "Generalized hypercube and hyperbus structures for a computer network," *IEEE Transactions on Computers*, vol. 33, pp. 323–333, apr 1984.

[64] A.-L. Barabási *et al.*, *Network science*. Cambridge university press, 2016.

[65] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Scientific american*, vol. 288, no. 5, pp. 60–69, 2003.

[66] R. I. Dunbar, "Neocortex size as a constraint on group size in primates," *Journal of human evolution*, vol. 22, no. 6, pp. 469–493, 1992.

[67] R. Dunbar, *How many friends does one person need?: Dunbar's number and other evolutionary quirks*. Faber & Faber, 2010.

[68] A. Hernando, D. Villuendas, C. Vesperinas, M. Abad, and A. Plastino, "Unravelling the size distribution of social groups with information theory in complex networks," *The European Physical Journal B*, vol. 76, no. 1, pp. 87–97, 2010.

[69] H. Ebel, L.-I. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," *Physical review E*, vol. 66, no. 3, p. 035103, 2002.

[70] D. Centola, "The spread of behavior in an online social network experiment," *science*, vol. 329, no. 5996, pp. 1194–1197, 2010.

[71] B. Doerr, M. Fouz, and T. Friedrich, "Why rumors spread so quickly in social networks," *Communications of the ACM*, vol. 55, no. 6, pp. 70–75, 2012.

[72] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," *arXiv preprint arXiv:1111.4503*, 2011.

[73] F. Karinthy, "Chain-links," *Everything is different*, 1929.

[74] S. Milgram, "The small world problem," *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.

[75] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, p. 440, 1998.

[76] A. D. Broido and A. Clauset, "Scale-free networks are rare," *Nature communications*, vol. 10, no. 1, pp. 1–10, 2019.

[77] P. Holme and B. J. Kim, "Growing scale-free networks with tunable clustering," *Physical review E*, vol. 65, no. 2, p. 026107, 2002.

[78] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[79] H. Lu, M. Halappanavar, and A. Kalyanaraman, "Parallel heuristics for scalable community detection," *Parallel Computing*, vol. 47, pp. 19–37, 2015.

[80] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[81] O. Goldreich, "Computational complexity: a conceptual perspective," *ACM Sigact News*, vol. 39, no. 3, pp. 35–39, 2008.

[82] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, "On modularity clustering," *IEEE transactions on knowledge and data engineering*, vol. 20, no. 2, pp. 172–188, 2007.

[83] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, oct 2008.

[84] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," in *2010 international conference on advances in social networks analysis and mining*, pp. 176–183, IEEE, 2010.

[85] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conference* (G. Varoquaux, T. Vaught, and J. Millman, eds.), (Pasadena, CA USA), pp. 11 – 15, 2008.

[86] M. Wackernagel and W. Rees, *Our ecological footprint: reducing human impact on the earth*, vol. 9. New society publishers, 1998.

[87] M. Wackernagel and B. Beyers, *Ecological Footprint: Managing our biocapacity budget*. New Society Publishers, 2019.

[88] M. Jalili, "Social power and opinion formation in complex networks," *Physica A: Statistical mechanics and its applications*, vol. 392, no. 4, pp. 959–966, 2013.

[89] S. Zhang, Z. Zhang, Y. Wu, M. Yan, and Y. Li, "Strategy preference promotes cooperation in spatial evolutionary games," *Physica*

A: Statistical Mechanics and its Applications, vol. 514, pp. 181–188, 2019.

[90] M. Perc, J. J. Jordan, D. G. Rand, Z. Wang, S. Boccaletti, and A. Szolnoki, "Statistical physics of human cooperation," *Physics Reports*, vol. 687, pp. 1–51, 2017.

[91] S. Bratsberg, "Modelling behaviour change dynamics on social networks towards climate change mitigation," Master's thesis, Norwegian University of Science and Technology, 8 2019.

[92] S. Khamis, L. Ang, and R. Welling, "Self-branding, 'micro-celebrity' and the rise of social media influencers," *Celebrity Studies*, vol. 8, no. 2, pp. 191–208, 2017.

[93] K. Freberg, K. Graham, K. McGaughey, and L. A. Freberg, "Who are the social media influencers? a study of public perceptions of personality," *Public Relations Review*, vol. 37, no. 1, pp. 90 – 92, 2011.

[94] A. Tavoni, M. Schlüter, and S. Levin, "The survival of the conformist: social pressure and renewable resource management," *Journal of theoretical biology*, vol. 299, pp. 152–161, 2012.

[95] M. Romenskyy, V. Spaiser, T. Ihle, and V. Lobaskin, "Polarized ukraine 2014: opinion and territorial split demonstrated with the bounded confidence xy model, parametrized by twitter data," *Royal Society open science*, vol. 5, no. 8, p. 171935, 2018.

[96] A. Baronchelli, "The emergence of consensus: a primer," *Royal Society open science*, vol. 5, no. 2, p. 172189, 2018.

[97] K. Sekimoto, "Kinetic characterization of heat bath and the energetics of thermal ratchet models," *Journal of the Physical Society of Japan*, vol. 66, no. 5, pp. 1234–1237, 1997.

[98] K. Sekimoto, "Langevin Equation and Thermodynamics," *Progress of Theoretical Physics Supplement*, vol. 130, pp. 17–27, 01 1998.

[99] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, "Probability of second law violations in shearing steady states," *Phys. Rev. Lett.*, vol. 71, pp. 2401–2404, Oct 1993.

[100] D. J. Evans and D. J. Searles, "Equilibrium microstates which generate second law violating steady states," *Phys. Rev. E*, vol. 50, pp. 1645–1648, Aug 1994.

[101] D. J. Evans and D. J. Searles, "The fluctuation theorem," *Advances in Physics*, vol. 51, no. 7, pp. 1529–1585, 2002.

[102] G. E. Crooks, *Excursions in Statistical Dynamics*. PhD thesis, University of California, Berkeley, 1999.

[103] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Phys. Rev. E*, vol. 61, pp. 2361–2366, Mar 2000.

[104] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, pp. 2690–2693, Apr 1997.

[105] C. Jarzynski, "Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale," *Annu. Rev. Condens. Matter Phys.*, vol. 2, no. 1, pp. 329–351, 2011.

[106] F. Naughton, "Coding and cats: What is this md thing anyway?," 2016.

[107] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method," *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 1011–1021, 1992.

[108] G. Hummer and A. Szabo, "Free energy reconstruction from nonequilibrium single-molecule pulling experiments," *Proceedings of the National Academy of Sciences*, vol. 98, no. 7, pp. 3658–3661, 2001.

[109] S. B. Howell, *Handbook of CCD astronomy*, vol. 5. Cambridge University Press, 2006.

[110] C. W. Groetsch, "Integral equations of the first kind, inverse problems and regularization: a crash course," *Journal of Physics: Conference Series*, vol. 73, p. 012001, jun 2007.

[111] É. Picard, "Sur un théorème général relatif aux équations intégrales de première espèce et sur quelques problèmes de physique mathématique," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 29, pp. 79–97, Dec 1910.

[112] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal*, vol. 79, p. 745, June 1974.

[113] W. H. Richardson, "Bayesian-based iterative method of image restoration∗," *J. Opt. Soc. Am.*, vol. 62, pp. 55–59, Jan 1972.

[114] M. K. Khan, S. Morigi, L. Reichel, and F. Sgallari, "Iterative methods of richardson-lucy-type for image deblurring," *Numerical Mathematics: Theory, Methods and Applications*, vol. 6, no. 1, p. 262–275, 2013.

[115] M. K. Singh, U. S. Tiwary, and Y.-H. Kim, "An adaptively accelerated lucy-richardson method for image deblurring," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 365021, Dec 2007.

[116] R. L. White, "Image restoration using the damped richardson-lucy method," *ASP Conference Series*, vol. 2198, 1994.

[117] H. Risken, "Fokker-planck equation," in *The Fokker-Planck Equation*, pp. 63–95, Springer, 1996.

[118] L. Sjögren, "Lecture notes on statistical physics: Chapter 7," 2014.

[119] S. Otsubo, S. K. Manikandan, T. Sagawa, and S. Krishnamurthy, "Estimating entropy production along a single non-equilibrium trajectory," *arXiv preprint arXiv:2010.03852*, 2020.