Computational Methods to Study Strongly-Binding Polymer-Colloid Systems

Computational Methods to Study Strongly-Binding Polymer-Colloid Systems

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam op gezag van de Rector Magnificus prof. dr. D.C. van den Boom ten overstaan van een door het college voor promoties ingestelde commissie, in het openbaar te verdedigen in de Agnietenkapel op dinsdag 9 december 2008, te 14:00 uur.

door

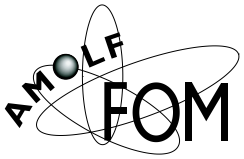
Behnaz Bozorgui

geboren te Teheran, Iran

Promotor: Prof. Dr. D. Frenkel

Overige leden: Prof. dr. Nienhuis
Prof. Dr. P. Bolhuis
Prof. Dr. W. Briels
Prof. Dr. M. Dogterom
Prof. Dr. C.N. Likos
dr. C. Lowe

Faculteit: Natuurwetenschappen, Wiskunde en Informatica



The work described in this thesis was performed at the FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ, Amsterdam, The Netherlands. The work is part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (FOM) and was made possible by financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Subject headings: / computer simulation / DNA-coated colloids / telechelics / phase transitions / pair-potentials / motor proteins

To My Mother, Batool.

The work in this thesis is based on the following publications

Chapter :2

Lattice-based Monte Carlo method for telechelic chain molecules.
B. Bozorgui and D. Frenkel.
Phys. Rev. E. **75**, 036708 (2007).

Chapter :3

Free-Energy-Based Method of Step Size Detection of Processive Molecular Motor.
B. Bozorgui, K. Shundyak, and D. Frenkel.
In Preparation.

Chapter :4

A finite-cluster phase in DNA-coated colloids.
T. Schmatko, B. Bozorgui, N. Geerts, D. Frenkel, E. Eiser, W. C. K. Poon.
Soft Matter **3**, 703-706 (2007).

Chapter 5:

Liquid-vapor transition driven by bond disorder.
B. Bozorgui and D. Frenkel.
Phys. Rev. Lett. **101**, 045701 (2008).

Chapter 5:

Liquid-vapor transition in Coated Colloid at Finite Temperature.
B. Bozorgui et al.
In Prepration

Other publication not in this Thesis:
Unexpected relaxation dynamics of self-avoiding polymer in cylindrical confinement.
A. Arnold, B. Bozorgui, D. Frenkel, B. Y. Ha, S. Jun.
J. Chem. Phys. **127**, 164903 (2007).

Contents

1 General introduction **1**

1.1 Simulations to mimic nature. 1

1.1.1 Synopsis 5

1.2 Useful terms 6

2 A lattice-based Monte-Carlo for telechelics **9**

2.1 Introduction 9

2.2 Grand-Canonical method 10

2.3 Detailed Balance 14

2.4 Efficiency 15

2.4.1 Wall-wall effective pair potential 15

2.4.2 Tethered polymers 18

2.5 Appendix A: Wormhole moves 22

2.6 Appendix B: splitting insertion arguments 23

3 A Step detection method for molecular motors **25**

3.1 Introduction 25

3.2 Analyzing artificial data 28

3.3 Moment propagation in matrix representation 32

3.4 Reference curves 35

3.5 Sensitivity 40

4 A Coarse-grained model for DNA coated colloids **47**

4.1 Introduction 47

4.2 Long DNA experiments 49

4.3 The model 52

4.4 Pair-potentials for grafted, sticky polymers 53

4.4.1 Colloid - polymer interaction at zero density 57

4.5 Reaction moves 60

4.5.1 Hybridization equilibrium of DNA 60
4.5.2 Reaction algorithm 63
4.6 Appendix A : Thermodynamics of DNA base-pairing interactions . 64

5 Liquid - vapor transition for DNA coated colloids 67
5.1 Introduction 67
5.2 Very high binding energy regime 68
5.3 Finite binding strength and fixed grafting points 76
5.4 Appendix A: Reaction switches 82
5.5 Appendix B: Cluster moves 83
5.6 Appendix C: Quaternions 83

Bibliography 87

Summary 92

Samenvatting 96

Acknowledgments 100

1 General introduction

1.1 Simulations to mimic nature.

The subject of this thesis is soft and biological matter. Soft matter covers a wide range of materials such as colloids, polymers, and surfactants[1], that on one hand are very complex because they consist of large numbers of atoms yet, at the same time are amenable to systematic experimental studies [1]. An important feature of many soft-matter systems is that the individual building blocks are so large that they can be imaged individually by optical microscopy, something that is usually not possible for simple atomic or molecular systems. As the name suggests, soft materials are easily deformable. The building blocks in soft materials are small enough to move appreciable on the time scale of an experiment under the influence of thermal fluctuations. These building blocks are thus able to explore configuration space and, as a consequence, the static properties of these materials can be described using the tools of equilibrium statistical mechanics.

An important feature of the phase behavior of many soft materials is the important role that entropy plays in determining the relative stability of different phases. In fact, already since the work of Onsager [2] it is known that entropic effects alone can cause colloidal rods to undergo a transition from the disordered "isotropic" phase to the orientationally ordered "nematic" phase. The early computer simulations of Alder and Wood [3, 4] showed that hard, spherical particles can undergo a transition from a disordered liquid state to an ordered crystalline state only because above some density the entropy of the crystal becomes larger than that of the liquid.

As the building blocks of soft materials are themselves composite objects, they can be designed such that they can self-assemble into unconventional phases that are rarely, if ever, observed in simple molecular systems. The ability to form very specific structures by self assembly is particularly for soft materials consisting of biomolecules. In the cell environment fluctuations due to thermal noise are significant and the specific interactions between biomolecules allow living organisms to

[1] *Concise definitions of "technical" terms, such as colloids, polymers and surfactants are given in section 1.2*

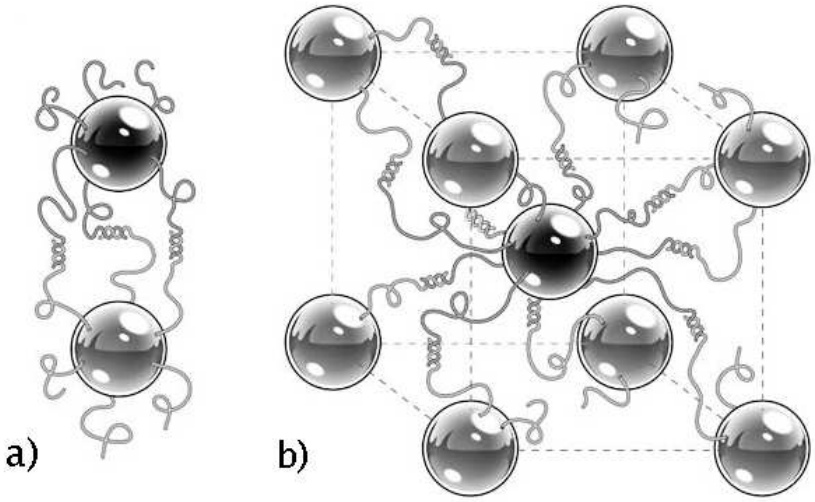


Figure 1.1: The schematic representation of two-type nano-scale gold particles coated with DNA strands. a) Dark particles (A) are coated with strands with ss DNA terminal parts which has complementary sequence to those grafted to light particles (B). When A and B are close the particles bind due to specific binding interactions between complementary ends. b) The self-assembly of these particles with CsCl structure that has been achieved in experiments of Refs. [5, 6].

construct intricate composite structures by self assembly. One of the reasons why biomolecules can form such complex structures is that the interaction between different molecules can be highly specific: a given molecule will therefore only bind to its putative partner and not to the myriad of other molecules in the cytosol. Yet, whilst the interaction between biomolecules is highly specific, it is usually not irreversible. To give a specific example: The four basic building blocks of DNA, i.e. the bases adenine (A), cytosine (C), guanine (G), and thymine (T), placed along the backbone of ssDNA strand, have a very specific interaction among themselves. The interaction between two complementary sequences of single-stranded DNA is highly specific and quite strong. However, by changing the physical conditions (e.g. changing the temperature or the salt concentration) the double-stranded DNA can be made to dissociate. This combination of specific interactions and reversibility makes biomolecular linkers, in particular those based on DNA, very interesting candidates for the design of novel "inanimate" self-assembling soft materials.

The highly specific nature of binding in double-stranded (ds) DNA offers many possibilities for the use of this biomolecule in micro- and nano-technology. For ex-

ample, colloidal particles coated with distinct species of short single-stranded (ss) DNA can bind specifically with each other via suitable DNA 'linkers' [7, 8, 9, 10]. These "suitable" linkers are themselves ssDNA strands added to the solution. Such DNA-coated colloids may offer a route to novel colloidal crystals [8, 10]. To make colloidal particles crystallize is an interesting problem in its general form, since it is one of the promising ways to control the structure and therefore properties of materials. This is not a new phenomenon for experimentalist: Without using any coating spacers or linkers, many particles do make ordered structure (as an example when the surrounding solution evaporates).

The "specific" DNA-mediated interactions may be made stronger than the "non-specific" interactions that normally dominate aggregation and self-assembly in colloidal systems. Recent experimental studies on gold nano-particles [5, 6] have successfully used short DNA strands to "direct" the self-assembly of particles into a non-close packed crystalline structure. Fig. 1.1 schematically shows the structure of these coated-colloid building blocks. Nevertheless, experiments that aim to achieve DNA-controlled self-assembly are very challenging. "Non-specific" and uncontrollable interactions remain one of the main challenges in controlling the structure of self-assemblies, and our understanding of the factors that control DNA-mediated interactions between colloids is still far from complete.

Computer simulations provide a powerful tool to probe the factors that determine the self assembly of complex bio-mimetic materials. The present thesis focuses on the development of computational tools to study the interaction between, and phase behavior of, colloidal particles that can bind to each other via linker molecules.

As the term "computer simulation" already suggests, our approach will be to investigate the behavior of bio-molecular systems by "mimicking" nature using computational tools. Here we face two conflicting requirements: on the one hand, our model description should be sufficiently realistic to reproduce the essential features of the system under consideration yet, at the same time, it should not be so detailed as to make numerical simulations prohibitively expensive. To this end, we must leave the atomic-level description of the components in the system under study and use a more coarse-grained approach where, for instance, colloids are not modeled as an assembly of billions of atoms but as simple, structureless objects (e.g. spheres or rods) that have an effective interaction that reproduces the interaction that we would have obtained if we would have taken all atoms into account explicitly. In addition, the atomistic description of the solvent is usually replaced by one where the solvent is described as a structureless, continuous medium. The key challenge in formulating a coarse-grained model is to construct accurate effective interactions between the

coarse-grained objects.

Once we have an adequate, coarse-grained model we need a numerical technique that allows us to compute the equilibrium properties of the system under consideration. The "work-horses" of computer simulation are, on the one hand, Molecular Dynamics (MD) and, on the other, Monte Carlo (MC) simulations (see e.g. [11]). In spirit, the Molecular Dynamics method is most closely related to experiment: average properties of a many-body system are computed by following the natural time evolution of a system for a sufficiently long time to obtain accurate averages. In such a simulation, the (Newtonian) equations of motion of all particles are solved approximately by discretizing time. At every time step, the force (and, if applicable, torque) on each particle is computed and this information is then used to update the velocities of all particles. Similarly, the positions of all particles are updated using the information about their instantaneous velocities. The MD technique is particularly useful if one is interested in the time-dependent behavior of a molecular system. However, for coarse-grained, soft materials, the motion of the (coarse-grained) particles is usually not described by Newton's equation of motion because the motion of coarse-grained particles through a continuous solvent is diffusive, rather than ballistic. Alternative algorithms exist to model the stochastic motion of particles in a hydrodynamic medium. However, as the focus of the present work is on equilibrium properties, I refer the reader for details to the relevant literature [12, 13, 14]. In the present thesis, I will make use of Monte Carlo simulations to study the structural properties of biomolecular systems. In a Monte Carlo simulation, we do not attempt to reproduce the natural dynamics of the system under consideration. Rather, the MC method performs an importance-weighted random sampling of the various configurations of the system, in such a way that configurations are visited with a frequency proportional to their Boltzmann weight. With this algorithm it is possible to compute equilibrium properties of a many-body system. However, the actual "dynamics" of the system is not related to the natural equations of motion. MC simulations are useful to study the static properties of systems that have no natural dynamics, but also to study systems where the natural dynamics is very slow, under conditions where MC sampling is still efficient. An example is a dense polymer solutions where the natural dynamics may be very slow, yet MC sampling based on the artificial insertion/deletion of entire chains in the system may be quite fast. Another advantage of MC simulations is that there is an unlimited reservoir of possible "unphysical" moves that can be designed to deal with specific conditions. It should, however, be stressed that a correct MC algorithm has to satisfy certain rigorous criteria that ensure that a correct Boltzmann sampling of configurations is achieved. Within these constraints, it is often possible

to find clever ways to "bias" the sampling of configurations in such a way that the efficiency of the simulation is enhanced [11] (See also section 1.2).

1.1.1 Synopsis

This thesis focuses on the use of simulations to study the behavior of colloidal particles that form multiple "bonds" via chain molecules that can either bind reversibly to the surface of the colloids ("telechelics") or via grafted polymers (in particular DNA) that can bind specifically to complementary polymers bound to other colloids. In chapter 2 we consider the development of suitable algorithms to simulate telechelic polymers that can bind strongly to a (colloidal) surface. It turns out that the combination of high binding strength on the surface and excluded volume interactions between the polymers makes "brute force" calculations difficult. For this situation, we propose a novel algorithm that combines the so-called "moment-propagation" to sample non-self-avoiding polymers with the configuration bias method to deal with self-avoidance effects. In chapter 2 we use a coarse-grained model to represent polymers as self-avoiding walks on lattice. We show that our approach is particularly useful to simulate dense systems of polymers with functionalized end groups. We compare the efficiency of the new scheme with a commonly used configurational bias MC method. We find that no single method is preferred under all conditions. We indicate the regime where the present approach is the method of choice.

Interestingly, it turns out that some of the methodology developed in chapter 2 can be applied in a completely different part of biophysics: we shall argue that the moment-propagation method offers a novel, and uniquely sensitive, method to identify underlying steps in noisy experimental data on the motion of molecular motors along linear "tracks" (e.g. kinesin on microtubules). This "detour" is the subject of chapter 3. The algorithm that we propose in chapter 3 is based on the numerical estimate of the "free energy" of an ensemble of directed random walks of different step sizes. We argue that this method allows us to detect the underlying step size of a molecular motor under "noisy" conditions where existing methods fail.

Chapters 4 and 5 then contain a description of the coarse-grained model that we use to simulate systems of colloidal particles coated with long DNA strands with "sticky" ends. In particular, chapter 4 describes a Grand Canonical GC simulation technique to study the self-assembly of mixtures of colloids that are coated with "sticky" DNA strands. Using a coarse-grained model to represent DNA strand as "soft" spheres, we study the effective interactions between the colloids. We develop a method that allows us to relate the binding properties of DNA-coated colloids to the (known) equilibrium constants for dimerization in solutions of the isolated "sticky" ssDNA

strands.

In chapter 5 we use the GC method mentioned above to simulate an equimolar mixture of hard colloids coated with "long" polymers with sticky ends. The sticky ends are "complementary" ssDNA strands that can selectively bind to each other. Under conditions where the complementary ends are strongly associated we observe a first order vapor-liquid transition from a dilute gas of colloidal dimers to a dense, liquid-like phase. This transition is driven exclusively by the increase in entropy associated with bond disorder. We show that the transition persists as we switch to low but finite temperatures.

1.2 Useful terms

Certain terms will be used quite time and again in this thesis. It is therefore useful to provide the reader with a (non-rigorous) working definition of the most frequently used terms.

Colloids are particles with a size between a few nanometers and a few microns. Because of their size they are too large to be described explicitly in terms of their detailed atomic structure. Yet they are small enough to be subject to thermal fluctuations. Colloids come in many shapes and chemical compositions. In fact, many familiar substances, from food products, paint and cosmetics to blood are colloidal suspensions. Part of the interest in colloids derives from the fact that they can be viewed as scaled-up models of atoms or molecules. In particular, they may exhibit the same phases as atoms and molecules: colloidal vapor, liquid , crystal and even liquid-crystalline phases have been observed. In addition, colloids may form glasses and gels. As colloids are much larger than atoms and move much more slowly they can often be studied by relatively simple optical microscopy.

Polymers are long chain molecules that are built from a repeat unit called a monomer. On a local scale, the properties of these chains are characterized by the structure of its monomers, but on a longer length scale, only the number, the effective size and the flexibility of the building blocks matter, as well as topology of the chain (e.g. linear or star like). Due to the large number of internal degrees of freedom, the statistical properties of polymers can often be estimated using the tools of statistical mechanics [15]. DNA is an example of a natural polymer: in its standard form, it is a double helix consisting of two single-stranded chains that consist of a sugar/phosphate backbone to which four characteristic "bases" may be connected:

adenine (A), cytosine (C), guanine (G) and thymine (T). Each of these bases can bind to a specific complementary base on the opposing strand of the double helix: C to G (via three hydrogen bonds) and A to T (via two hydrogen bonds) (see e.g. [16]).

Telechelics are polymers that have functionalized end groups that can bind to specific targets. In this thesis we are specifically interested in telechelics with two functional end groups that can link to surfaces and to double-strand (ds) DNA that is "functionalised" by having a single-stranded (ss) end sequence. The ssDNA acts as "sticky" endgroup that can bind to a complementary ssDNA group on another dsDNA chain. The important characteristic of these DNA-based telechelics is that the binding of these sticky ends is very specific as it depends on the sequence of ssDNA.

Molecular motors are biological machines that perform a plethora of mechanical and chemical tasks in a living cell. Some motors act as transporters of "cargo" through the cell. These motors walk on fairly rigid protein filaments such as microtubules. Other motors may drive the relative motion of (actin) fibers, for instance in muscle action. Molecular motors use the chemical energy released during the hydrolysis of ATP to perform mechanical work.

Grand Canonical (GC) ensemble is an ensemble of configurations for a system in which volume, temperature T, and chemical potential μ, are fixed and the number of particles is allowed to fluctuate. The system is considered to be in equilibrium with a large reservoir, i.e. $T = T_{\rm reservoir}$ and $\mu = \mu_{\rm reservoir}$. In GC simulations, normal Monte Carlo moves are combined with particle exchange moves between the system under consideration and a fictitious reservoir at constant chemical potential [11]. In this thesis we use the GC ensemble to study interacting polymers (chapter 2), and DNA-coated colloids (chapter 4 and 5). Detailed descriptions of simulation techniques are presented in each chapter.

Detailed Balance (DB) is a sufficient (but not necessary) condition that ensures that a Monte Carlo simulation will not drive a system out of thermal equilibrium. It is based on the assumption that, in thermal equilibrium, the number of transitions from any state i to any other state j should, on average, be balanced exactly by the number of transitions from j to i. Throughout the thesis we will employ MC algorithms that satisfy Detailed Balance.

Biased sampling is a generic term that refers to a wide range of methods that change (usually enhance) the acceptance probability of Monte Carlo trial moves by biasing the probability with which different moves are attempted. Several biased MC schemes are discussed in Ref. [11]. In chapter 2 we introduce a specific biased scheme to sample configurations of dense systems of telechelic polymers.

2 A lattice-based Monte-Carlo for telechelics

2.1 Introduction

Polymer simulations are often time-consuming because the natural dynamics of polymers is slow. For this reason, many Monte Carlo methods have been developed that sample the configurational space of polymeric systems by carrying out efficient "unphysical" trial moves, i.e. moves that do not correspond to the slow natural motion of polymeric chains, yet do preserve the Boltzmann distribution (see, e.g. Reptation [17, 18], Pivot moves [19], Configurational-Biased Monte-Carlo(CBMC) [11], Connectivity-altering [20], Hyper-Parallel Tempering [21], Wormhole moves [22].) In spite of the great variety in computational schemes, there are situations where polymer simulations are still very inefficient. This is particularly true in the case of polymers that can bind to heterogeneous surfaces. It is, for instance, time consuming to compute the free energy of a system of polymers that can either be non-adsorbing or form bridges and loops between surfaces.

In this chapter, we describe a Grand-Canonical Monte-Carlo technique that is particularly efficient for simulations of "Telechelic" polymers that interact strongly with surfaces. The basic idea is to use the statistics of non-self-avoiding walks to grow self-avoiding walks (SAW) on a lattice. Whilst there exists no "cheap" method to enumerate the number of SAW chain conformations in a given system, efficient algorithms exist to count the number of polymer conformations that correspond to ideal and non-reversible (NR) random walks. NR walks are random walks that exclude 180-degree reversal: i.e. retracing the last step is excluded. We can count the number of NR walks using the so-called moment-propagation (MP) scheme [23]. To be more precise, during a trial move we enumerate all NR walks for one polymer, while keeping the positions of all other polymers fixed. The exact enumeration of NR walks implies that if there is only a single acceptable conformation for the additional polymer, it will be included in the enumeration. However, some acceptable NR configurations may not be self-avoiding and will therefore be rejected at a later stage.

This pre-screening of acceptable configurations greatly enhances the success rate of particle-insertion moves. In fact, as we will show, the present method becomes much more efficient than CBMC at high densities. In the following, we explain how we can incorporate the number of NR walks in a lattice-based CBMC scheme and use this scheme to grow self-avoiding chains. We will demonstrate that the MC acceptance rule satisfies detailed balance. Finally, we show a number of results for non-adsorbing and for telechelic chains, on the one hand to validate the present method and, on the other, to quantify its relative performance with respect to other existing schemes.

2.2 Grand-Canonical method

Consider a regular lattice with coordination number z. We assign a set of z numbers to each lattice site. Each number represents a weight which is related to one of the neighbors. Let i and j be two neighboring sites on the lattice. $\omega'(i; j \rightarrow i; l)$ is defined as the Boltzmann-weighted number of non-reversible walks of length l, which end at site i along the bond $j \rightarrow i$. Hence $\omega'(i; j \rightarrow i; l)$ can be viewed as the partition sum of a single NR chain that has its end segment along the link $i - j$. As NR walks cannot retrace their last step, it is necessary to label not simply the number of walks arriving at site i, but also the lattice site last visited. For instance, if we wish to calculate $\omega'(i; j \rightarrow i; l + 1)$ knowing the corresponding weights for walks of length l we should exclude those walks that went form i to j in the previous step

$$\omega'(i, j \rightarrow i; l + 1) = \left(\sum_{j' \in \langle j \rangle'} \omega'(j, j' \rightarrow j; l) \right) \omega(i, 0). \tag{2.1}$$

where $\langle j \rangle$ is defined as the set of z nearest neighbors of site j, and by $\langle j \rangle'$ we mean the conformation-dependent set which excludes the last step (site i in the above case). $\omega(i, 0)$ is defined as the Boltzmann weight due to the interaction of a monomer at site i with other polymers or with external interactions, such as walls. This Boltzmann weight is a measure of the accessibility of site i for a monomeric unit (chain of length 0). As an example, for hard monomers, $\omega(i, 0) = 0$ when the site is occupied and 1 when it is empty. It is then sufficient to know $\omega(i, 0)$ for all lattice sites i to calculate all the weights for any length.

The weights ω' determine the likelihood of growing a non-reversible walk with a fixed length l coming from a specified direction. Using these weights in the insertion algorithm will introduce a bias that facilitates the finding of acceptable conformations

for the polymer to be inserted. It would, of course, have been possible (in fact, easier) to calculate the number of ideal random walks and their corresponding weights. However, as very few ideal walks are completely NR, the subsequent step in the algorithm would reject most of the walks thus generated, because at some points, they retrace their steps. For this reason, the use of NR weights is more efficient.

Insertion algorithm Having computed all weights $\omega'(i_1, j \rightarrow i_1; n)$ for $n = 0, 1, \cdots l$, we can now start inserting the self-avoiding chain, using these weights as a bias. Lattice sites from which there emerge large numbers of NR walks will be favored over those that spawn none, or only few. In what follows, we will assume that non-bonded monomers of the same chain cannot occupy the same lattice site, but do not interact otherwise. If that is the case, self-avoidance effects only come into play upon growing the 4th polymer segment (at least, on a simple cubic lattice). However, it is straightforward to generalize the present approach to the case where intra-chain interactions are longer ranged. We will also assume that all the polymers are of the same length. We generate a trial conformation of a chain with fixed length l following these steps:

1. The first monomer i_1 is selected with probability

$$p_1 = \frac{\sum_{j \in \langle i_1 \rangle} \omega'(i_1, j \rightarrow i_1; l)}{Q_{eff}}$$

   The sum in the numerator is a sum over all z neighbors of site i_1 and it gives the total number of non-reversible walks of length l terminating at site i_1 from different directions. Q_{eff} is the total number of NR walks of length l in the system of M lattice sites:

$$Q_{eff} = \sum_{i=1}^{M} \sum_{j \in \langle i \rangle} \omega'(i_1, j \rightarrow i_1; l). \tag{2.2}$$

   The above expression simplifies to $Mz(z-1)^{l-1}$ for an empty lattice containing M sites.

2. The second monomer is selected from one of the z neighbors of i_1, called i_2, with a probability

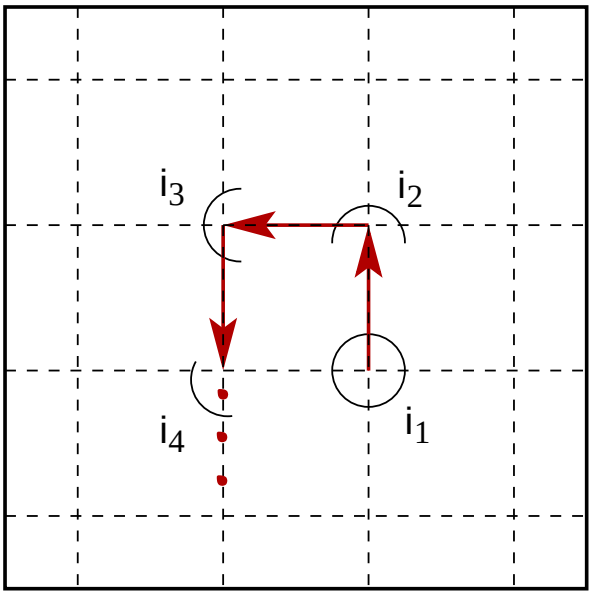


Figure 2.1: Schematic representation of moment propagation algorithm. The arrows show the trace of polymer chain as it has been building from the first monomer i_1. The curved lines indicate the possible candidates for the next monomer to be chosen. As it has been shown the curved line at i_4 excludes the position of monomer i_1. See the algorithm in the text.

$$p_2 = \frac{\omega'(i_1, i_2 \to i_1; l)}{\sum_{j \in \langle i_1 \rangle} \omega'(i_1, j \to i_1; l)}$$

3. The third monomer is selected from one of the neighbors of i_2 with probability

$$p_3 = \frac{\omega'(i_2, i_3 \to i_2; l-1)}{\sum_{j \in \langle i_2 \rangle'} \omega'(i_2, j \to i_2; l-1)}.$$

4. The fourth monomer is selected from one of the neighbors of i_3 with probability

$$p_4 = \frac{\omega'(i_3, i_4 \to i_3; l-2)}{\sum_{j \in \langle i_3 \rangle'} \omega'(i_3, j \to i_3; l-2)}.$$

5. Starting with the fifth monomer we have to take into account the self-avoidance. The fifth monomer is chosen with the following probability,

$$p_5 = \frac{\omega'(i_4, i_5 \to i_4; l-3)\exp[-\beta u_{int}(i_5)]}{\sum_{j\in\langle i_4\rangle'}\omega'(i_4, j \to i_4; l-3)\exp([-\beta u_{int}(j)]}$$

The "internal" energy $u_{int}(i_s)$ accounts for the interactions with all monomers 1 to $s-1$, that have been already grown.

6. Step 5 is repeated until the whole chain is grown (unless the growth process terminated in a dead-end, due to intra-chain interactions). The probability of choosing the s-th monomer equals to

$$p_s = \frac{\omega'(i_{s-1}, i_s \to i_{s-1}, l+2-s)\exp[-\beta u_{int}(i_s)]}{\sum_{j\in\langle i_{s-1}\rangle'}\omega'(i_2, j \to i_2, l+2-s)\exp[-\beta u_{int}(j)]}$$

Once we have grown the whole chain, the move is accepted with the following probability,

$$\begin{aligned}&\mathrm{acc}(n_p \to n_p+1) = \\ &\min\left\{1, \frac{\exp(\beta\mu)Q_{eff}\prod_{l=1}^{N}\langle\exp[-\beta u_{int}(i_l)]\rangle}{n_p+1}\right\}.\end{aligned} \tag{2.3}$$

In the above equation, the terms in angular brackets denote the weighted averages of Boltzmann factors of the internal potential along the chain, defined as

$$\begin{aligned}&\langle\exp[-\beta u_s(i_s)]\rangle \equiv \\ &\frac{\sum_{j\in\langle i_{s-1}\rangle'}\omega'(i_2, j \to i_2, l+2-s)\exp[-\beta u_{int}(j)]}{\sum_{j\in\langle i_{s-1}\rangle'}\omega'(i_2, j \to i_2, l+2-s)}.\end{aligned} \tag{2.4}$$

To clarify the meaning of the acceptance rule, we decompose it in three different parts. The first part, $\exp[\beta\mu]/(n_p+1)$ comes from the chemical potential or, equivalently, the average polymer density in the "osmotic reservoir" in contact with the system. We know this part before growing a chain or calculating any weight. The second part is Q_{eff} which comes from the NR guiding weights. To know this part we have to calculate NR weights before each trial of exchanging a new chain. However

this part does not depend on the specific configuration that we grow. The only part in the acceptance argument that depends on the configurations is the product term. If we had ignored self-avoidance affects in the NR chain, the terms in Eqn. 2.4 would have been equal to unity and the acceptance probability would have been determined only by Q_{eff}.

The main difference between the new biased scheme and CBMC is the presence of non-local guiding weights ω'. In general, in a CBMC scheme a trial conformation will be generated using local conformation-dependent weights. In the new scheme these weights carry long range information about acceptable conformations of the whole system. This facilitates the growing process and will increase the acceptance rate. In the next section, we show that the proposed algorithm satisfies detailed balance.

Now consider the situation in which a chain is randomly selected to be completely removed. First, we remove the chain and re-calculate the guiding weights ω'. Then we reconstruct the same old configuration monomer by monomer. At each step we calculate the weighted average of u_{int} which is defined in equation 2.4. After the whole chain has been retraced we will remove it with the probability:

$$\begin{aligned} &\text{acc}(n_p + 1 \to n_p) = \\ &\min\left\{1, \frac{n_p \exp[-\beta\mu]}{Q_{eff}\prod_{l=1}^{N}\langle\exp[-\beta u_{int}(i_l)]\rangle}\right\}. \end{aligned} \tag{2.5}$$

We considered several ways of improving upon the above algorithm. Since we know some parts of the acceptance argument before trying to grow a configuration, we could split the rejection criterion in two or three parts (See Appendix 2.6). This might increase the computational efficiency as we can reject a "doomed" trial configuration at an early stage. However, it turns out that splitting the acceptance procedure in parts is not helpful for dense systems – and it is precisely for these systems that the present approach is most competitive. For this reason, we used the acceptance rule in the form given in Eqn. 2.3 and Eqn. 2.5.

2.3 Detailed Balance

The acceptance rule in Eqn. 2.5 should satisfy Detailed Balance (DB). Let us define $\alpha(n_p \to n_p + 1)$ as the probability of generating a new configuration, and acc$(n_p \to n_p{+}1)$ as the probability of accepting it and let $\alpha(n_p{+}1 \to n_p)$ and acc$(n_p{+}1 \to n_p)$ denote the corresponding quantities for the reverse move. DB implies that:

$$\frac{P_B(n_p)\alpha(n_p \to n_p+1)}{P_B(n_{p+1})\alpha(n_p+1 \to n_p)} = \frac{\text{acc}(n_p+1 \to n_p)}{\text{acc}(n_p \to n_p+1)}, \tag{2.6}$$

where $P_B(n_p)$ ($P_B(n_{p+1})$) denotes the Boltzmann weight of the configuration with n_p (n_{p+1}) polymers. Looking back at the insertion algorithm we can derive $\alpha(n_p \to n_p+1)$ by calculating the probability of the new configuration.

$$\begin{aligned}
&\alpha(n_p \to n_p+1) = \prod_{s=1}^{l} p_s = \\
&\frac{\sum_{j\in\langle i_1\rangle} \omega'(i_1, j \to i_1; l)}{Q_{eff}} \\
&\times \frac{\omega'(i_1, i_2 \to i_1; l)\exp(-\beta u_{int}(i_2))}{\sum_{j\in\langle i_1\rangle} \omega'(i_1, j \to i_1; l)\exp(-\beta u_{int}(j))} \\
&\times \frac{\omega'(i_2, i_3 \to i_2, l-1)\exp(-\beta u_{int}(i_3))}{\sum_{j\in\langle i_2\rangle'} \omega'(i_2, j \to i_2, l-1)\exp(-\beta u_{int}(j))} \\
&\times ... \\
&\times \frac{\omega'(i_{s-1}, i_s \to i_{s-1}; l+2-s)\exp(-\beta u_{int}(i_s))}{\sum_{j\in\langle i_{s-1}\rangle'} \omega'(i_{s-1}, j \to i_{s-1}; l+2-s)\exp(-\beta u_{int}(j))} ...
\end{aligned}$$

We can rewrite the above equation in a simple form using Eqn. 2.1 and Eqn. 2.4:

$$\alpha(n_p \to n_p+1) = \frac{\exp(-\beta U_{int})}{Q_{eff}\prod_{s=1}^{l+1}\langle\exp(-\beta u_{int}(i_s))\rangle}.$$

where $U_{int} = \sum_s u_{int}(i_s)$ denotes the total internal energy of the chain. By inserting $\alpha(n_p \to n_p+1)$ in Eqn. 2.6 we recover the acceptance rule that we have used in the insertion algorithm.

2.4 Efficiency

2.4.1 Wall-wall effective pair potential

To validate the present method, we first applied it to compute the depletion interaction due to non-adsorbing, self-avoiding polymers between hard walls. This system provides a convenient yet non-trivial test of our algorithm, as Bolhuis et al. [24] and Tuinier et al. [25] have calculated the depletion interaction of SAW polymers between

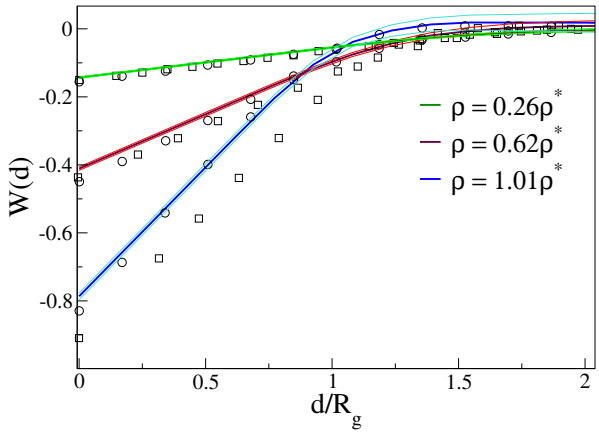


Figure 2.2: Depletion interaction for athermal, self-avoiding lattice polymers between unstructured hard walls. The strength of the depletion interaction is expressed in units $k_{\mathrm{B}}T/R_g^2$. The three sets (lines) correspond to three different densities: 0.26, 0.62 and 1.01 ρ^*, where ρ^* denotes the overlap concentration (see text). The parallel shadowed lines indicate the error bars. The corresponding data points of Ref. [24] and Ref. [25] are indicated as squares and circles.

hard walls using Configurational Bias MC, over a wide range of polymer densities. The depletion attraction can be computed by determining the grand-partition-function Ξ, at fixed chemical potential μ. For each separation between two walls a thermodynamic integration relates the grand-partition function Ξ, to the average number of particles in the system $\langle N(\mu')\rangle$,

$$k_B T \ln \Xi(\mu, d) \quad = \quad \int_{-\infty}^{\mu} \langle N(\mu')\rangle d\mu'. \tag{2.7}$$

The depletion interaction follows by subtracting the bulk contribution to the grand potential and the contribution due to two non-interacting surfaces. In what follows, we denote the radius of gyration of the polymers by R_g. The overlap concentration ρ^* will be defined as $1/v_p$, where $v_p = 4/3\pi R_g^3$ is the effective volume of a single polymer coil. Systems with concentrations well below ρ^* are considered to be dilute. Here, we report results for three densities and compare the results with those reported in Ref. [24] and [25].

We carried out grand-canonical MC simulations for self-avoiding polymers of

length $l = 100$. The polymers were simulated on a cubic lattice between two hard walls. Several simulation were carried with different spacing between the walls. In the direction parallel to the walls, we employed periodic boundary conditions. The lateral size of the periodic box was 50×50 lattice sites. Hard-core repulsions were defined to exclude conformations where two monomers occupy the same lattice site. For the Monte Carlo moves we used the "wormhole method" of Houdayer [22] (See Appendix 2.5). The MP method described in previous section was used for polymer insertion and deletion. We calculated the average number of particles as a function of chemical potential. We started from a very dilute regime (corresponding to $\mu = -\infty$) to get a correct estimation of Ξ. When the surface separation is large, $kT \ln \Xi$ contains only the bulk contribution (that can be computed for a system without walls) and the constant contribution due to two non-interacting walls, $2\gamma_w A$, where A is the surface area of the hard wall. Hence, by comparing $kT \ln \Xi$ for a bulk system and a system with the same number of lattice sites, but enclosed between two hard walls, we immediately obtain γ_w. Fig. 2.2 shows the results of these simulations for three different concentrations. Within the regime where scaling arguments are valid, the depth and range of the potential only depend on the radius of gyration and concentration. As expected, the depth of the potential increases with increasing polymer density. In view of the large difference between the results of refs. [24] and [25], our results appear consistent with both (unfortunately, neither paper quotes error bars). Our data are closest to those of Ref. [25]. We note that, in order to reproduce the bulk densities of Ref. [24] and Ref. [25] for a chain with length 100 we assumed $R_g \approx 6.476$ as was calculated by Nickel [26]. Our error estimates in Fig. 2.2 have been calculated by taking into account the statistical errors in integration, fitting and subtracting.

Compared to the CBMC method, MP has a higher insertion rate per MC trial move; in the above case, it is some 100 times higher. However, the CBMC method is much less time consuming. We performed the same simulation with the CBMC method to compare the efficiency and found CBMC to be about 5000 times faster than MP in the above case. With both methods higher densities can be simulated. For chains of length $L = 100$ densities up to $3\rho^*$ are feasible. To conclude, with the MP method we can reproduce the earlier CBMC results. However, for the specific problem of non-adsorbing polymers between hard walls, CBMC is clearly the method of choice.

Although MP fails to bring an improvement for systems with only hard core interactions, it turns out to be more useful in problems with different kinds of interactions. MP is far more efficient in identifying rare configurations with very low energy (i.e. with large Boltzmann weight). A good example is a system of polymers with

strongly interacting end groups. Such polymers can bind to each other or to specific binding sites on surfaces. A case in point is the binding of DNA chains with single-stranded end segments, to colloids coated with complementary single-stranded (ss) DNA. Such polymers can form bridges between colloids or form loops on single surfaces. The interactions between surfaces are the combination of bridging attraction, steric repulsion of the loops or dangling polymers, and depletion of the free linkers. In order to compute the effective interaction free energies between the surfaces, it is crucial, yet difficult, to sample well equilibrated systems. Here the advantage of the present moment-propagation scheme over CBMC becomes clear.

2.4.2 Tethered polymers

In order to test the efficiency of MP on the above-mentioned class of problems, we performed simulations on systems with end-associating polymers between two walls. Apart from the end-group interactions, the system is the same as the one described in the previous section. Each polymer is considered as a SAW chain which in this case can attach to the surfaces with its terminal groups, due to a strong, short-ranged attraction. In what follows, we will consider the specific case that the two walls are identical. This implies that polymers can either form loops by adsorbing both their ends on one wall, or form bridges by attaching to both surfaces. In addition, the polymers may have one or two end groups unbound. The range of the binding interaction is set to one lattice spacing, meaning that the binding sites are placed next to the hard wall. The surface grafting density ρ_s measures the fraction of binding sites at each layer. The binding energy ϵ_b is defined as the depth of the attractive potential. Other parameters such as the bulk density ρ and the separation between the hard walls affect the computational efficiency. Most of the existing methods fail to equilibrate systems at high densities. Also finding bridged conformations is increasingly difficult at higher densities, especially when the separation between the surfaces is large compared to R_g (because then the most favorable links are highly stretched).

Fig. 2.3 shows a typical example of the difference in performance between CBMC and MP. The results shown in this figure correspond to a situation where the grafting density $\rho_s = 0.02$, which corresponds to 50 sites per wall in our case. We chose a relatively large binding energy: $\epsilon_b = 10 k_{\mathrm{B}} T$. Such large values are not unphysical: for instance, the hybridization free energy with which DNA strands bind to ssDNA coated colloids is in this range. For instance, the hybridization of ssDNA of 14 bases is in the range of $25 - 30 k_{\mathrm{B}} T$ [27]. In our simulation every chain has two interacting ends, i.e. the energy of a bound configuration is $2\epsilon_b = 20 k_{\mathrm{B}} T$. The bulk

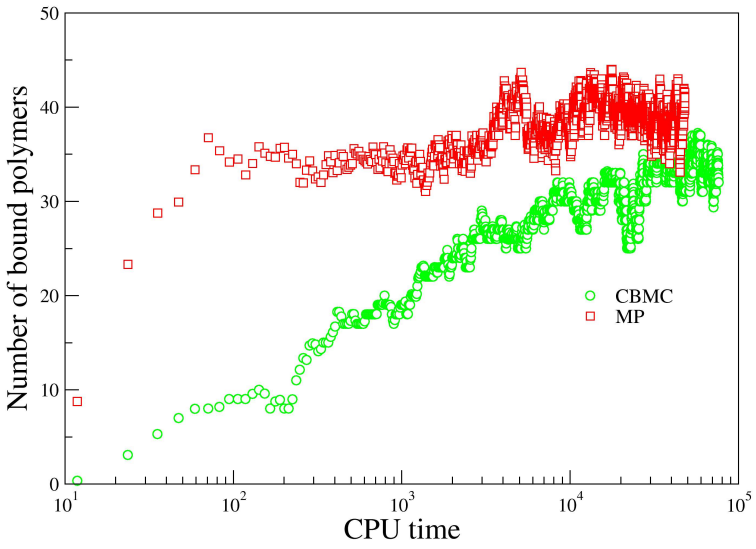


Figure 2.3: Rate of equilibration of the number of bound (bridge or loop) polymers between two walls as computed using the MP scheme (squares) and the CBMC scheme (circles). The horizontal axis indicates the required CPU time. The spacing between the two walls is 8 lattice units. The volume fraction of monomers is 0.23. The fraction of binding sites on the surface, $\rho_s = 0.02$. The binding energy, $\epsilon_b = 10 k_{\mathrm{B}} T$.

monomer volume fraction, $\rho_{mon} = 0.23$. As can be seen from the figure, the CPU time needed for equilibration is much less when using the MP method than with the CBMC method. In fact, CBMC seems to fail to equilibrate the system, even for the longest runs. We note that the two methods approach equilibrium in a very different way. Starting from very dilute initial conditions, the MP scheme first grows bound configurations. In most cases, these bound configurations reach their equilibrium value even faster than those that are free. The reason why the MP scheme favors configurations with higher Boltzmann weight in the selection process is that the Boltzmann-weighted numbers ω' (see Eqn. 2.1) provide information about accessible configurations. By contrast, CBMC initially fills the system with dangling and

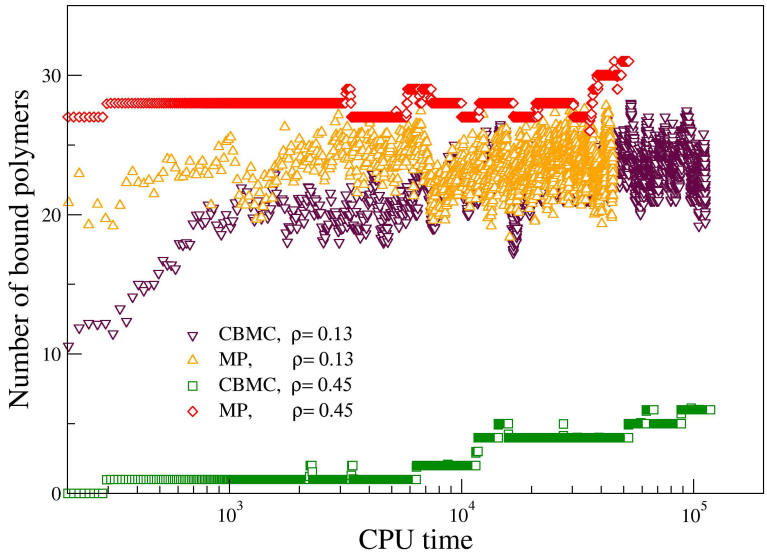


Figure 2.4: Density dependence of the equilibration rate of the MP and CBMC schemes. The figure shows the number of bound (bridge or loop) polymers between two walls as a function of CPU time. As in figure 2.3, the spacing between the two walls is 8 lattice units. The fraction of binding sites on the surface, $\rho_s = 0.02$ and the binding energy, $\epsilon_b = 10 k_{\mathrm{B}} T$. Results are shown for a high monomer concentration (45%) and a low monomer concentration (13%). The inset shows the density dependence of the number of bound polymers.

free polymers. This makes it subsequently more difficult to reach the equilibrium density of bound configurations.

To check the efficiency of our method we performed simulations for three different grafting densities $\rho_s = 0.02$, 0.2, and 1.00, and different binding energies $\epsilon_b/(k_{\mathrm{B}}T) = 5$, 10. We also studied the system both in a narrow gap $d = 3$, and in larger distances like $d = 8$, and we tried different bulk densities.

On the basis of these simulations, we reach the following tentative conclusions:

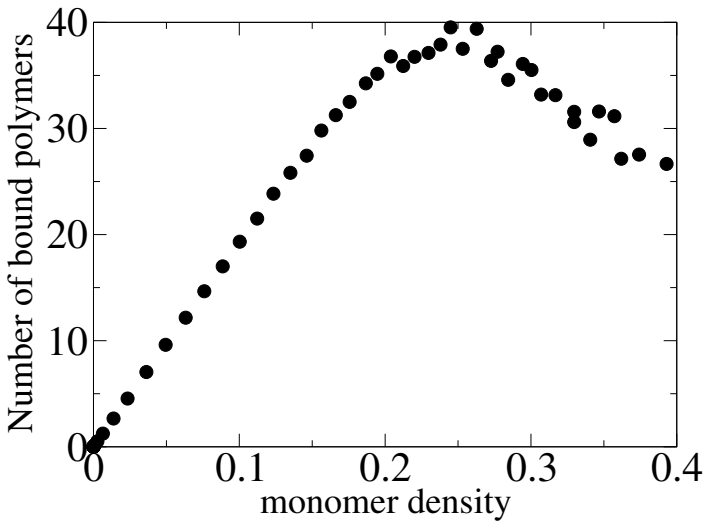


Figure 2.5: The density dependence of the number of bound polymers.

At high grafting densities the MP scheme is less useful. This comes about because in equilibrium all polymers will be bound. Then all configurations have the same energy and there is no need for a scheme (such as MP) that will identify rare configurations that have a much higher Boltzmann weight than most others. As one might expect, both methods become inefficient at high densities and large wall-wall separations. However, the results show that increasing polymer density affects CBMC much more adversely than MP. We compared the two methods at low and high densities in fig. 2.4. At high densities, CBMC fails to find bound chains, while MP is much closer to the equilibrated system. The MP method is therefore the method of choice in cases where there is a large heterogeneity in binding energies and in systems where the accessible volume per polymer is small. Both situations are relevant for the study of selective binding of biomolecules to specific substrates. Interestingly, the results shown in figs. 2.3 and 2.4 suggest that the number of bound polymers is not a monotonic function of density. It first increases with increasing density and then decreases (See fig. 2.5). We can see this effect with the novel approach presented in this chapter, but not with CBMC. Still, we should caution that, even with the MP method, it is difficult to obtain high statistical accuracy at high polymer densities.

2.5 Appendix A: Wormhole moves

The wormhole method proposed by Houdayer [22] includes special reptation moves that allows a polymer to perform an unphysical Monte Carlo move whereby it "reptates" segment-by-segment from one part of space to another position that, in general, will not be near its original position. This feature of the algorithm explains the name "wormhole move". During a move (that may consist of a large number of reptation moves) the polymer is in an unphysical state where it is split in two parts that are connected respectively to the "entrance" and the "exit" of the wormhole. A wormhole move is only complete when the complete chain is reassembled again, either at the exit or at the entrance of the wormhole. Configurations of the system during wormhole moves are unphysical and are not sampled.

In practice, the algorithm works as follows:

Consider a linear chain on lattice [1] with N monomers labeled i_1 to i_N.

1. Choose one of the terminal monomers i_1 or i_N at random, with probability 0.5. For the sake of the argument, we will assume that i_1 was thus chosen.

2. Move the chosen monomer to a random trial position that is chosen from a uniform distribution over the simulation box. If this position is already occupied, the trial move is rejected (and this wormhole move is terminated). However, if the new position is unoccupied (or more generally, if the reptation move is acccepted), proceed to the next step.

3. This move breaks the existing bond between i_1 and i_2 and, by analogy to the conventional reptation scheme, a "bond" is now formed between i_1 and i_N. However, this is only a virtual bond: it is introduced to maintain the analogy with normal reptation. In the new situation, the end points of the new polymer configuration are i_1 (now at the end of the wormhole) and i_2 that is still connected to the entrance of the wormhole (located at i_N).

4. Randomly select a monomer at either endpoint and attempt a trial move that will reconnect it to the other endpoint. Use the normal (e.g. Metropolis) rule to accept or reject this trial move.

5. If, after this move, the polymer is still straddling the wormhole (i.e. i_1 and i_N are connected via the wormhole) repeat step 4. Otherwise, the polymer has

[1] *Although we consider lattice polymers here, the method also works for off-lattice polymers*

been reassembled (at either end of the wormhole) and the wormhole move has been completed.

It is important to stress that, even if the wormhole move "fails" in the sense that the polymer ends up at the same side of the wormhole where it started, the state of the polymer will, in general, have changed because the chain will have reassembled in a different conformation at the entrance of the wormhole. This speeds up relaxation. In addition, the computing effort for the wormhole scheme appears to scale favorably with polymer size for long polymers (namely as N^2 rather than as $\exp(cN)$).

2.6 Appendix B: splitting insertion arguments

The wormhole algorithm discussed in appendix A is a particular example of a composite MC move where an acception-rejection criterion is applied to every sub-step. In the case of Grand-Canonical Monte Carlo simulations of polymers, we can also decompose the trial moves into sub-steps that may be accepted/rejected individually. However, the difference with the wormhole move is that, in this case, a rejection of any substep implies the rejection of the entire move. Let us consider the typical condition on the acceptance probability of a particle-insertion/removal trial move that follows from the condition of detailed balance.

$$\frac{acc(n_p \to n_p + 1)}{acc(n_p + 1 \to n_p)} = P. \tag{2.8}$$

Now suppose that the ratio P can be expressed as the product of several (non-negative) factors that can be computed individually and sequentially. In the present case, we assume that P can be expressed as $P = ABC$, with A, B and $C \geq 0$. Then we can satisfy Eqn. 2.8 if we use the following acceptance rules:

$$\begin{aligned} acc(n_p \to n_p + 1) &= min\{1, A\} min\{1, B\} min\{1, C\} \\ acc(n_p + 1 \to n_p) &= min\{1, \frac{1}{A}\} min\{1, \frac{1}{B}\} min\{1, \frac{1}{C}\} \end{aligned} \tag{2.9}$$

In the special case that we consider in the present chapter where B and C are always less than 1, the acceptance rule becomes

$$\begin{aligned} acc(n_p \to n_p + 1) &= min\{1, A\} BC \\ acc(n_p + 1 \to n_p) &= min\{1, \frac{1}{A}\} \end{aligned}$$

In our case, in Eqn. 2.3, we choose A, B, and C as

$$\begin{aligned} A &= \frac{\exp[\beta\mu]Q_{id}}{n_p+1} \\ B &= \frac{Q_{eff}}{Q_{id}} \\ C &= \prod_{l=1}^{N}\langle\exp[-\beta u_{int}(i_l)]\rangle, \end{aligned} \tag{2.10}$$

where Q_{id} is the partition function of a chain on an empty lattice. Since the effective partition function in solution is less than Q_{id}, the second term is always less than one. The same holds for the third term $< \exp[-\beta u_l(i_l)] >$, as $\exp[-\beta u_l(i_l)]$ can take only the values 0 or 1; therefore the average can never exceed one. The acceptance rule for insertion or removal of a self avoiding polymer will then become:

$$\begin{aligned} acc(n_p \rightarrow n_p+1) &= min\{1, \frac{\exp(\beta\mu)V}{n_p+1}\}\frac{Q_{eff}}{Q_{empty}}\prod_{l=1}^{N} < \exp[-\beta u_l(i_l)] > \\ acc(n_p+1 \rightarrow n_p) &= min\{1, \frac{n_p+1}{\exp(\beta\mu)V}\} \end{aligned}$$

Note that if we use the above expressions, we need not calculate any weights for removing a particle. We therefore, will be able to save time, since calculating the weights is the most time consuming part in the moves.

3 A Step detection method for molecular motors

3.1 Introduction

As mentioned in the section 1.2, molecular motors are biological machines that drive active motion in living cells. They perform mechanical work using ATP (or a similar energy-rich molecule) as a fuel. The energy that is needed for the power-stroke of the molecular motor is released upon hydrolysis of the ATP. There exists a wide range of molecular motors. Some of these transport a cargo from one point in the cell to the other, others drive the relative motion of cytoskeletal filaments, a task that is important for processes such as cell motility or muscle action. Different molecular motors "walk" on different filaments: for example, the myosins that are responsible for muscle action can move along actin filaments whilst kinesins are the motors that transport cargo along microtubules.

An important characteristic of many of these motor proteins is that they move effectively unidirectionally. The direction of motion is determined by the type of motor and by the polarity of the filament on which they move. The hydrolysis of ATP drives a sequence of configurational changes in the motor with the net effect that the motor proteins make one step of fixed length for every ATP molecule that they burn. The size of these steps is typically of the order of several nanometers and the net speed of the motors is of the order of $\mu m/sec$. In most cases, the speed of the motor depends on the force exerted by the load, but the step-size is constant.[1]

Experiments that provide information about the step-size and stepping rate of molecular motors under different conditions provide crucial information about the mechanism of the motor action and, indirectly, about the function of the motor. Experiments that can track the position of individual molecules have made it possible to track a single motor protein as it moves through the cell. However, such experiments are notoriously difficult because molecular motors operate in an environment where fluctuations due to thermal noise are significant. Therefore, optical-microscopy ex-

[1] *Dynein, for example, is an exception since its step-size depends on the external load.*

periments that follow the motion of a molecular motor along a track do not only detect the underlying motion of the motor but also the thermal fluctuation of the optical probe around this trajectory. As an illustration, Fig. 3.1 shows a sample trace of the motion of kinesin as observed in the experiments of Ref. [28]. It should be stressed that kinesin represents a favorable case as it typically travels for hundreds of cycles along a microtubule without detaching. In contrast, other motor proteins, such as myosin II, tend not to act alone but as part of a large number of similar motors: every individual motor does detaches from the (actin) fiber after every single power stroke. Such behavior makes it even harder to get the desired information from the noisy trace of the motor. Yet, as can be seen in Fig. 3.1, even for the case of kinesin, the level of noise is significant compared to the step-size of the motor. For this reason, much attention has been paid to the development of analysis techniques to extract information about the motion of molecular motors from the noisy traces that are obtained in experiment. Below we briefly review the existing approaches and then we present a different approach that makes it possible to extract information about step-size and stepping frequency under conditions where the existing techniques are either inaccurate or fail completely.

To extract the step-size and step frequency from noisy traces, such as the one shown in Fig. 3.1, several step detection methods have been proposed in the literature [29, 30, 31, 32]. We briefly describe some of these methods in the last section of this chapter. Here we simply refer to the method by their key characteristic: "t-test" in Ref. [29], χ^2-minimization for Ref. [30], velocity thresholding in Ref. [31] and a Markovian-based approach in Ref. [32]. A common feature of all these schemes is that their objective is to fit a *specific* "optimal" step pattern to the experimental data. The methods of Refs. [29, 30, 31] use different techniques to decompose the experimental data into pure noise and a noise-free stepping trace. Then, they extract information about the underlying trace from the reduced noise-free trace. The performance of all these fitting schemes strongly depends on the elimination of noise from noise-free trace and their performance rapidly deteriorates as the level of noise increases.

The key observation that underlies the approach that we propose here is the following: what we aim to extract is not the precise trajectory of the molecular motor, but the parameters (step-size(s) and step frequency) that are intrinsic properties of the molecular motor under the experimental conditions. There may be many step histories that fit the experimental data to within the experimental error and we argue that it is then pointless to select one trajectory as the "true" motion of the motor and to discard all others. Rather, we argue that is often better to extract the desired information

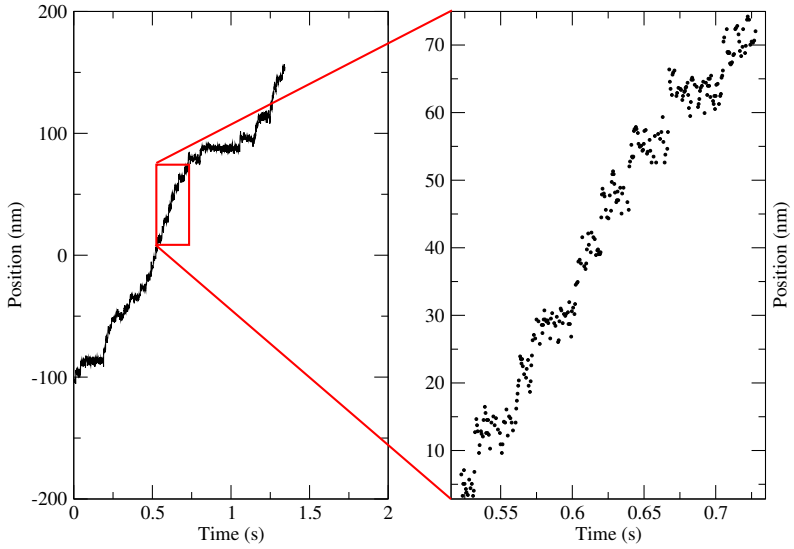


Figure 3.1: Trace of kinesin along microtubules that is observed with a quadrant detector under condition of very low ATP concentration. The step-size is reported to be 8nm [28]. The fixed directionality of kinesin is toward the plus end of microtubule. A detail of the experimental trace is shown on the right: it shows both the stepping behavior and the noisiness of data.

from the set of possible trajectories.

To illustrate why it may be advantageous to consider sets of trajectories, we consider a hypothetical situation where we have two sets of parameter values (e.g. two values for step-size) that yield an equally good "optimal" fit to the experimental data, yet differ in the *number* of good fits that are found. In this case a conventional method will consider both sets of parameter values equally acceptable. However, the number of good fits or, more precisely, the path integral of possible trajectories, is a better indicator of the quality of the fit. Naively, one might think that it is simply the number of acceptable trajectories that determines the goodness of fit. However, as we shall argue below, a slightly more sophisticated criterion must be used. The method based on the analysis of the properties of an "ensemble" of paths becomes more relevant at the high level of noise where the single fit result becomes inaccurate.

In essence, our approach is Bayesian in nature, as it considers the probability that a given set of parameters (e.g. step-size and jump frequency) are compatible with the

experimental data. In what follows we will exploit the analogy between stepping-motor trajectories and realizations of a directed random walk. The quality of a fit is related to the 'free energy' of the directed random walks in the landscape determined by the experimental data. We can define the Hamiltonian of a specific random walk $y(t;p)$ in the presence of experimental data $x(t)$ as:

$$H(y(t),\{p\}) \equiv \sum_t (x(t) - y(t;\{p\}))^2,$$

where $\{p\}$ denotes the set of parameters characterizing the random walk. If the statistical noise in the experimental data is denoted as σ^2, we can define an inverse 'temperature' $\beta \equiv 1/\sigma^2$. The "partition function" associated with the parameter set $\{p\}$ is then

$$Q(\{p\}) \equiv \sum_{y(t;\{p\})} e^{-\beta H(y(t),\{p\})}$$

As the number of trajectories that contribute to Q increases exponentially with the length of the trace, special techniques are needed to evaluate Q. In what follows, we will make use of "moment propagation", a technique that makes it possible to perform an exact, recursive enumeration of the weighted number of non-self-avoiding random walks [33, 34, 35]. In the context of the present example, we limit the discussion of moment propagation to the case of directed random walks. Moreover, we limit ourselves to motors that never step backwards. This restriction is not essential but it simplifies the subsequent discussion. However, we will show that the extension of the method to other cases is relatively straightforward.

3.2 Analyzing artificial data

In order to develop a reliable test for the approach that we propose, it is useful to work with artificial data, i.e. data that are generated by 1) assuming a particular step-size and jump frequency for the molecular motor 2) generating a specific trajectory for this motor with a probability that is determined by these parameters and then 3) add artificial Gaussian noise to the trajectory to generate a trace that can be compared to the experimental data. The reason why we perform tests on artificial data is that, in that case, we *know* the properties of the underlying random walk and the added noise and we can then unambiguously test how well our analysis method recovers the underlying information about the step-size from the noisy data. Fig. 3.2 shows such an artificially produced data set, which resembles molecular motor trace in position-time (x,t) space (similar to Fig. 3.1). Every point $x(t)$ represents the measured

displacement of the motor along one dimensional substrate at a certain time. We assume that the experimental data set originates from a homogeneous Markov process. By definition, a Markov process is a random memoryless process, i.e. the events that take place at any time depend only on the state of the system at that point in time and not explicitly on its earlier history.

The directional motion of a motor in time axis can be viewed as a "directed random walk" in $1 + 1$ dimensions. In the present case, these directions correspond to time t and position x. The random walker in this $1 + 1D$ space always goes forward in time. Such a random walk is called directed random walk (DRW). For simplicity, we consider the case where the process underlying the directed random walk only performs jumps in the $+x$-direction of a single, fixed size (e.g. kinesin [28]). In the language of DRW this situation corresponds to a random-walker that can move only forward in both the x and the t directions: a "unidirectional random walk" (URW).

Fig. 3.2 shows an example of an artificially generated set of N points $\vec{x} = (x_0, x_1, \ldots, x_{N-1})$. The data points correspond to the sum of a specific realization of a URW of step-size Δ_r (dashed curve) and the added noise. In the case of real experimental data, the observer has no a priori knowledge of the underlying step-size Δ_r and the aim of the analysis is to obtain an estimate of this quantity. The simplest case to consider is one where the stepping process is Markovian: At each time t the motor may either stay at the same position or may make a step of fixed size Δ_r where the probability to make a step does not depend on its previous history. For simplicity, we also assume that the hidden URW was sufficiently over-sampled upon acquisition of the experimental data; this means that at most one step can occur between two consecutive data points. This is not a crucial assumption and we will show later that the method can be extended to the case of under-sampled data. Without loss of generality, we measure time t in units of interval between the subsequent time frames, i.e. $t = \{0, 1, 2, \ldots N - 1\}$ with N the maximum measurement time, and we fix location of the first measured point at origin $(x_0 = 0)$[2].

For the set of artificial data in Fig. 3.2 the added noise to the underlying trace is a white Gaussian noise with distribution $G(0; \sigma)$, which means the distribution has zero mean and constant standard deviation σ. As we mentioned earlier, the presence of noise obscures the underlying URW. We assume that the magnitude of the noise can be determined independently in the experiments (for instance, by measuring the random, apparent displacements of a motor that is not stepping - for instance, for lack of fuel.)

[2] *To keep the notation simple, we drop the explicit dependence on the origin x_0. However, it should be borne in mind that, in practice, x_0 should also be varied.*

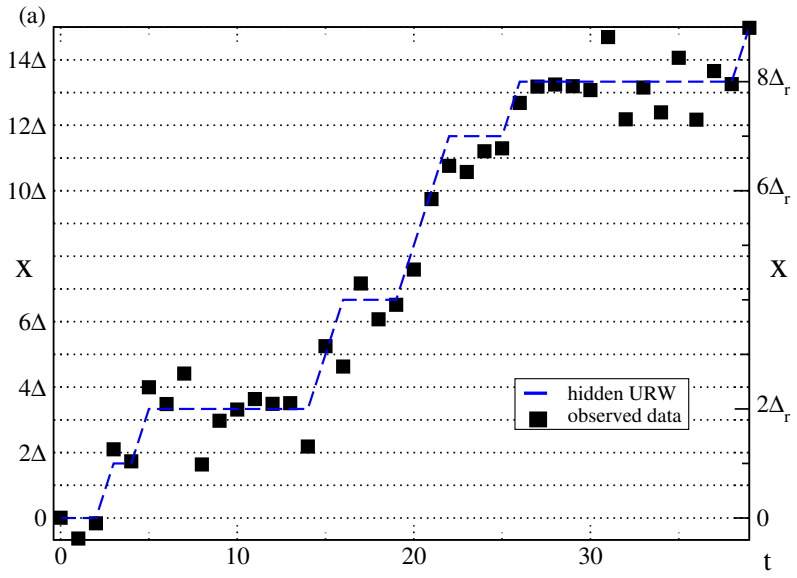


Figure 3.2: (a) Position-time trace $\vec{x} = (x_0, \ldots, x_t, \ldots x_{N-1})$, $t \in [0, N-1]$, of an artificially generated set of $N = 40$ points, which is a sum of a unidirectional random walk (dashed curve) of step-size Δ_r (see right-hand vertical axis) and white Gaussian noise $G(0, \sigma)$, $\sigma/\Delta_r = 0.5$. A priori, the observer does not know the true step-size Δ_r. In the text we consider calculations that compute the partition function of trial URW's for various step-sizes Δ. The left vertical axis indicates an example of the step-size of the uni-directional random walk (URW). Only when Δ_r equals a multiple of Δ will the horizontal stretches of the dashed curve fall in the horizontal (dotted) lines of the "probe grid".

The distance in time between subsequent data points defines the discretization Δt of the URW in the t direction. A random walk with step-size Δ, starting at the origin, should then only visit the grid points of a lattice with spacings Δt and Δ in respectively the time and the space directions. A random walk with added noise or with a step-size $\Delta_r \neq \Delta$ will, in general, not visit the grid points of the $\Delta t - \Delta$ lattice. The total number of URW's with length t and step-size Δ is denoted by Ω_Δ. Note that most of these walks will, in general, be far removed from the data points. In order to obtain a measure of the number of URW's with step-size Δ that are close to the experimental data, we have to give a weight to every URW that is determined

by its distance to the data points.

Now consider a particular trajectory $\vec{y}_\Delta$, $\vec{y}_\Delta = [y_0\Delta, y_1\Delta, \ldots, y_{N-1}\Delta]$.

How well this trajectory fits to the (synthetic) data depends on the positions and the size of the steps and, of course, to the magnitude of the statistical noise in the data. The deviation of a specific trajectory $\vec{y}_\Delta$ from the data points (denoted by $\vec{x}$) can be expressed as a 'Hamiltonian'

$$H(\vec{x} - \vec{y}_\Delta) \equiv \sum_t H_{t,\Delta} = \sum_t (x_t - y_t\Delta)^2. \tag{3.1}$$

The probability $p(\vec{x}|\vec{y}_\Delta)$ that the data points $\vec{x}$ are the result of an underlying URW $\vec{y}_\Delta$ plus Gaussian noise with variance σ is then proportional to a "Boltzmann weight" p defined as:

$$p(\vec{x}|\vec{y}_\Delta) \propto \exp(-\beta H(\vec{x} - \vec{y}_\Delta)) \ , \tag{3.2}$$

where the 'inverse temperature' β is defined as $\beta = 1/(2\sigma^2)$. In general, the underlying random walk will not start exactly at the position of the first data point. We therefore define Δ_0 as the displacement between the first data point and the trial random walk. The likelihood that the observed experimental data $\vec{x}$ originates from random walks on a lattice with parameters $\{\Delta, \Delta_0\}$ is determined by the "free energy"

$$f(\Delta, \Delta_0) = -\frac{1}{N} \ln Z(\Delta, \Delta_0), \tag{3.3}$$

$$Z(\Delta, \Delta_0) = \sum_{\{y\}} \exp(-\beta H(\vec{x} - \vec{y}_\Delta)), \tag{3.4}$$

where $Z(\Delta, \Delta_0)$ is the partition function of trajectories in ensemble Ω_Δ, and summation is carried over all possible $\{\vec{y}_\Delta\}$ and the standard factor $1/N$ makes $f(\Delta, \Delta_0)$ intensive.

The following argument may clarify the relation between the "likelihood" of one class of trajectories and the free energy: $H(\vec{x} - \vec{y}_\Delta)$ is smallest for trajectories that fit the data best. Therefore, an ensemble of URW's with a large number of good fits has a large "partition function" and hence a lower free energy. It would, however, be incorrect to assume that the quality of a fit is determined only by the free energy. To see this, consider a trial random walk with a very small step-size $\Delta \ll \Delta_r$. Most of these trial walks do not resemble the data set at all yet, because Δ is so small, there will be a very large number of trial walks that happen to be very close to the data. In Sec. 3.4 we will discuss how to resolve this ambiguity. However, for the time being, we concentrate on the calculation of the partition function Z by exact enumeration.

For all grids with different step-size Δ we need to calculate the sum in Eqn. 3.4 considering all possible URW's on the lattice that start from $x_0 = 0$. For each trajectory we calculate $H(\vec{x} - \vec{y}_\Delta)$ using Eqn. 3.1. Note, that σ as it is defined in Eqn. 3.2 determines fluctuations of x_t. Once we know the noise of the experimental data, we can calculate β and hence $\exp(-\beta(x_t - y_t\Delta)^2)$ for each point y_t on the grid and store it. Then for each trajectory y_Δ we calculate $H(\vec{x} - \vec{y}_\Delta)$. As the number of possible walks (Ω) grows exponentially with the number of data points (N), it is of course impossible to perform the summation over paths $\vec{y}$ explicitly except for very small N. Hence, it would seem that the computation of the sum in Eqn. 3.4 might be a numerically intractable problem. Fortunately, explicit summation is not needed. In the next section we show how a recursive algorithm, (moment propagation), allows us to perform the required summation exactly as an $\mathcal{O}(N)$ operation.

3.3 Moment propagation in matrix representation

In the previous chapter, we introduced the moment-propagation technique as a convenient tool to calculate partition sums of non-self-avoiding random walks on lattice without the need to store all the walks separately. We now apply the same technique (suitably generalized) to the enumeration of the "partition function" of directed random walks with a Hamiltonian that is determined by their distance from a given data set.

Let us define a vector of weights $\vec{w}(t) = [w_1(t), w_2(t), \ldots, w_N(t)]$ for every time t such that

$$\vec{w}(t+1) = \hat{B}(t)\hat{S}\vec{w}(t), \tag{3.5}$$

where $N \times N$ matrices $\hat{B}$ and $\hat{S}$ will shortly be introduced. As we assume that a motor can make at most one jump in one time step, the (hypothetical) maximum distance that a motor could move in N time steps of length Δt is $N \times \Delta$. We need therefore only consider URW's on an $N \times N$ grid (strictly speaking, only the lower triangular part). At a given time t, the data point that results from the experiment (or, in the present case, from the synthetic data) is denoted by x_t. Most likely, the underlying random walk will lie close to this data point. However, we consider all possible values of $y_t = \Delta_0 + j\Delta$. The further point y_t is removed from x_t, the lower is its Boltzmann weight. We introduce an $N \times N$ dimensional, diagonal matrix $\hat{B}(t)$ that contains the Boltzmann weights of deviations of all lattice nodes from x_t at time t:

$$B_{ij}(t) = \delta_{ij} \exp\left(-\beta(x_t - (\Delta_0 + j\Delta))^2\right), \tag{3.6}$$

where δ_{ij} denotes the Kronecker delta, $\{i, j\} \in [1, N]$ and $\Delta_0 + j\Delta$ is the position of lattice node (t, j) on the grid of vertical step-size Δ and offset Δ_0 from the origin. Note that typically only a few elements of B will differ significantly from zero. $\hat{B}$ can be computed explicitly as we know the experimental data $\vec{x} = (x_0, x_1, \ldots, x_{N-1})$ and σ, the variance of the noise. The propagation matrix $\hat{S}$ is a $N \times N$ Jordan matrix with eigenvalue 1, i.e.,

$$\hat{S} = \begin{vmatrix} 1 & 0 & 0 & 0 & \ldots \\ 1 & 1 & 0 & 0 & \ldots \\ 0 & 1 & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{vmatrix}. \tag{3.7}$$

The explicit representation of Eqn. 3.5 is:

$$w_i(t+1) = B_{ij}(t) S_{jk} w_k(t) \tag{3.8}$$

where the Einstein convention of summation over repeated indices is implied. Writing the Jordan matrix explicitly, and using the fact that $\hat{B}$ is diagonal, we obtain:

$$\begin{aligned} w_i(t+1) &= B_{ij}(t)(w_j(t) + w_{j-1}(t)) && (3.9) \\ &= B_{ii}(w_i(t) + w_{i-1}(t)), && (3.10) \end{aligned}$$

where $B_{ii} = \exp\left(-\beta(x_t - (\Delta_0 + i\Delta))^2\right)$ is the Boltzmann weight of node (t, i). We can now interpret weights $w_i(t)$ as the partition function of directed walks that end at node (t, i). The above equation implies that the total number of walks that end at node $(t+1, i)$ is the sum of two parts: 1) those walks that end in (t, i) and then continue to $(t+1, i)$ without a step in x direction, and 2) those walks that converge on node $(t, i-1)$ and then step to the node $(t+1, i)$. The partition function $Z(\Delta, \Delta_0)$ is then expressed as

$$Z(\Delta, \Delta_0) = \vec{w}(N+1)\hat{B}(N)\hat{S}\ldots\hat{B}(2)\hat{S}\hat{B}(1)\hat{S}\vec{w}(0). \tag{3.11}$$

Eqns. (3.5)-(3.11), together with the definition of $\vec{w}(t)$, initial condition $\vec{w}(0) = (\exp(-\beta\Delta_0), \ldots, 0, 0)$ and $\vec{w}(N+1) = (1, \ldots, 1)$ completely specify the moment-propagation algorithm.

As the Boltzmann weights B_{ij} decay quickly when the distance of the lattice point to the data point is significantly larger than σ, it is in practice sufficient to evaluate the

elements of $\hat{B}$ only in a band around x_t, for instance $x_t \pm 10\beta\Delta^2$. Another practical point to note is that Z is obtained by successive multiplications and additions. Due to the repeated multiplications, the propagated numbers can grow or shrink exponentially (in the present implementation, numbers tend to shrink). In order to prevent numerical underflows we keep track of the largest element at every propagation step, $w_{max}(t)$, where $w_{max}(t)$ is the maximum of the set $[w_1(t), w_2(t), \ldots, w_N(t)]$. This element is then rescaled to one and all other elements of $[w_1(t), w_2(t), \ldots, w_N(t)]$ are multiplied with the same scaling factor. The value of scaling factor (or, more precisely, its logarithm) is then used to compute the overall rescaling factor that must be taken into account when computing the end result of the moment-propagation algorithm. In this way, the algorithm can compute partition functions over a very wide "dynamic range".

Generalizations of the current method to the more complex symmetries of the underlying random walks can be easily constructed via modification of 'propagation' matrix $\hat{S}$ of Eqn. 3.7. For example, random walks with forward and backward steps correspond to $\hat{S}$ with 1's on its super- and sub-diagonal:

$$\hat{S} = \begin{vmatrix} 1 & 1 & 0 & 0 & \ldots \\ 1 & 1 & 1 & 0 & \ldots \\ 0 & 1 & 1 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{vmatrix}. \tag{3.12}$$

This class of random walks for example can represent stepping behavior of microtubules, where both growth and shrinkage can occur. Another example is URW's that may contain double steps in a single time interval. In that case, $\hat{S}$ has two bands of 1's below the diagonal. Such a formulation would be useful in cases where stepping is under-sampled:

$$\hat{S} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0\ldots \\ 1 & 1 & 0 & 0 & 0\ldots \\ 1 & 1 & 1 & 0 & 0\ldots \\ 0 & 1 & 1 & 1 & 0\ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{vmatrix}. \tag{3.13}$$

Another situation that occasionally has to be considered in practice is that the data set is incomplete, i.e. that at certain times data points are missing, or more generally, when the data are not equally spaced in time. In the case of missing data points, the matrix $\hat{B}(t)$ is replaced by the identity matrix $\hat{I}$.

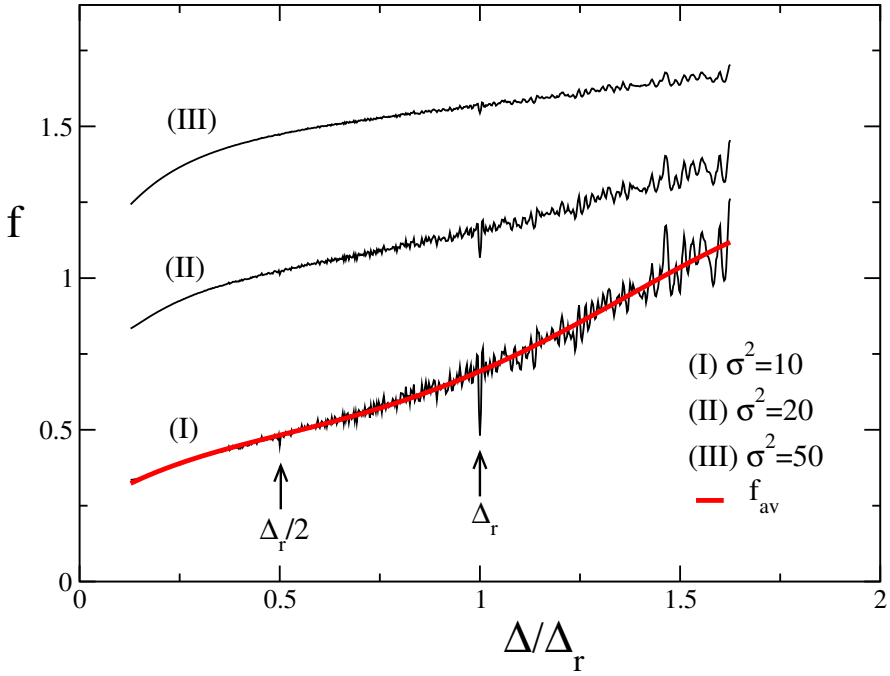


Figure 3.3: Free energy f as a function of the dimensionless step-size Δ/Δ_r for three independent artificially generated data sets with $\Delta_0 = 0$ and different noise levels $\sigma/\Delta_r = \sqrt{10}/8$ (I), $\sqrt{20}/8$ (II) and $\sqrt{50}/8$ (III). For clarity, the curves (II) and (III) were shifted by 0.25 in vertical direction. Arrows indicate peaks at the correct step-size Δ_r and $\Delta_r/2$ The grey line corresponds to the least-square polynomial fit f_{av}.

3.4 Reference curves

Fig. 3.3 shows the free energy $f(\Delta, \Delta_0)$ as a function of the dimensionless step-size Δ/Δ_r for a (synthetic) data set of $N = 4000$ points for three different values of σ the amplitude of statistical noise. For simplicity, we have assumed that we know the initial motor position in the underlying URW (in this case, $\Delta_0 = 0$). A pronounced and sharp local minimum in the free energy can be seen (at least, in the low-noise case) at $\Delta = \Delta_r$, i.e. at the correct step-size. As is to be expected, the minima become less pronounced as the noise level increases. For the highest noise level, the

minimum is no longer distinguishable by eye. Interestingly, a secondary minimum can be seen at $\Delta_r/2$, i.e. at half the real step-size. The presence of such a minimum is not surprising: we have not made any a priori assumption of the average time interval between successive steps. Hence, a trajectory with step-size Δ can be fairly well retraced by a trajectory with steps of $\Delta/2$, provided that steps always bunch in pairs around the position of the true steps in the original data. The fact that such subsidiary minima can exist makes it important to have a procedure that will allow us to distinguish between a physically meaningful fit and an "artificial" free-energy minimum. The more so as Fig. 3.3 clearly shows that the absolute minimum in the free energy is not related to the best choice of the step-size.

In order to extract information about the step-size from curves such as the ones shown in Fig. 3.3, we need the tools to assess whether or not a given feature in the free-energy curve is compatible with a particular step-size in the underlying random walk. A way to do this is to define a proper "reference curve". As we will shortly see even for the perfect matching between hidden underlying URW and the probing grids, i.e., $\Delta_r = \Delta$ and $\Delta_0 = 0$, our method generates some non-zero free energy. The value of this free energy is not constant and depends on Δ. Also it cannot be obtained by simple scaling.

We therefore define a "reference curve" $f_{ref}(\Delta, \Delta_0)$ as the free energy of a trace where the step-size of the hidden underlying URW is equal to the spacing of the probing grid. We argue (and we will shortly show) that information we get from the excess free energy $f_{ex} = f - f_{ref}$ is more likely to be generated by the stepping properties of the underlying walk, as f_{ref} has all the characteristics that comes from the noise in the data or the roughness of the grid. When the free energy of the experimental trace f is calculated on the "right" grid, the value of f should be close to f_{ref} and, therefore, f_{ex} will be close to zero.

For every step-size Δ, we then would like to calculate the value of the reference free energy $f_{ref}(\Delta, 0)$ on the "right" grid, i.e. a grid that perfectly matches the underlying data. We simply do this with the moment propagation method of Eqs. (3.5)-(3.11), using a 'reference' trace $\vec{R}(\Delta)$, which is a sum of the noise-free URWs of step-size Δ and Gaussian noise σ. This $\vec{R}(\Delta)$ should not be arbitrary, but should resemble the experimental data $\vec{x}$, e.g. in the least-square sense. For simplicity, in the present analysis we pick up an arbitrary $\vec{R}(\Delta)$ which starts at origin and ends up within $\pm\Delta$ of x_{N-1}, i.e., $x_{N-1} - \Delta \leq R_N(\Delta) \leq x_{N-1} + \Delta$. Fig. 3.4 shows the reference free energy $f_{ref}(\Delta, 0)$ for $\sigma/\Delta_r = \sqrt{10}/8$. Note, that the level of noise in this curve is almost constant, in contrast to the free energy profiles of Fig 3.3.

In order to use f_{ref} as a reference curve we need to calculate the average properties

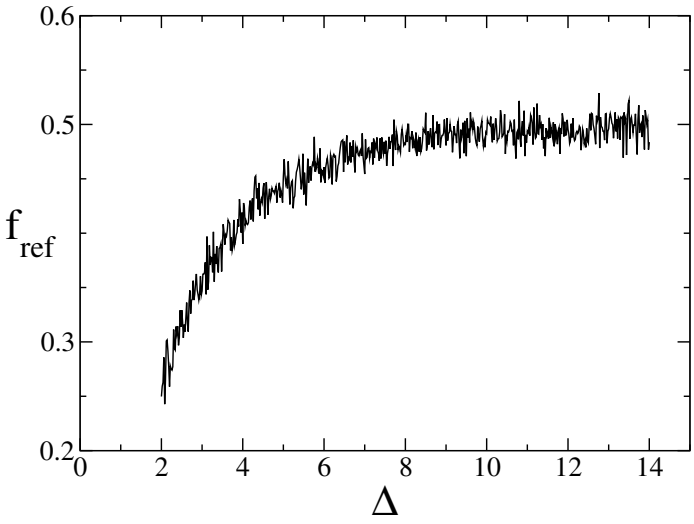


Figure 3.4: Reference free energy $f_{ref}(\Delta, 0)$ for a $\sigma = \sqrt{10}$ (see text). Note that the level of noise is almost constant in contrast to Fig. 3.3. Also see Fig. 3.5.

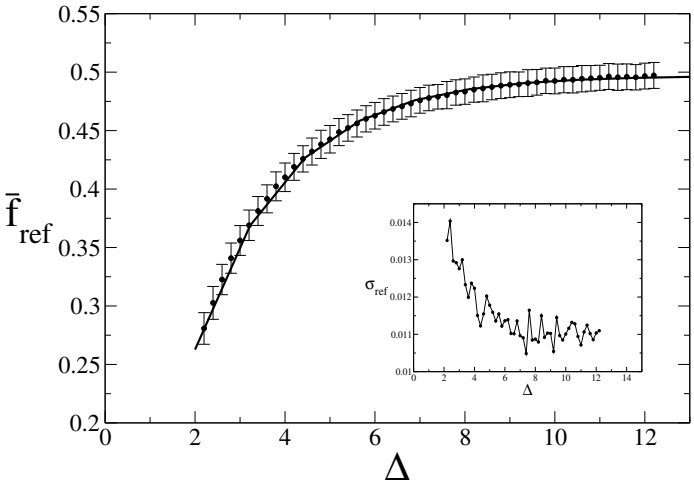


Figure 3.5: The average reference free energy $\bar{f}_{ref}(\Delta, 0)$ (data points) as a function of step-size Δ for $\sigma = \sqrt{10}$. The fitted dashed line is the least-square fit $a + b\exp(-cx)$, $a = 0.496 \pm 0.001$, $b = -0.628 \pm 0.001$, $c = -0.493 \pm 0.001$. The error-bars show the value of σ_{ref} calculated for each data point. The inset shows that σ_{ref} increases for the smaller step-sizes.

of it, namely the mean value $\bar{f}_{ref}$ and the standard deviation σ_{ref}. We obtain the average smooth curve $\bar{f}_{ref}(\Delta, 0)$ via averaging over ensemble of $\{\vec{R}(\Delta)\}$ each starts at origin and ends up within $\pm\Delta$ of x_{N-1}. All these traces have the same level of noise, while the position of steps in the underlying URW may differ from each other. Fig. 3.5 shows $\bar{f}_{ref}$ as a function of the step-size. $\bar{f}_{ref}$ increases at large step-sizes. For large step-sizes ($x \gg 1$) the only significant contribution to $f_{ref}(\Delta, 0)$ comes from thermal wandering around perfect matching trajectory and $\bar{f}_{ref}$ converges to 0.5. The error bars show the σ_{ref}, the noisiness of the free energy curve. It seems that the roughness of f_{ref} slightly decreases at larger step-sizes (See inset of Fig. 3.5).

In order to calculate f_{ex}, we approximate $\bar{f}_{ref}$ by a simple functional form:

$$\bar{f}_{ref}\,(x, 0) = a + b \exp\left(-cx\right). \tag{3.14}$$

that provides a good fit to the data. We use this fit to evaluate the excess free energy $f_{ex}(\Delta, 0) = f(\Delta, 0) - \bar{f}_{ref}(\Delta, 0)$ which is shown in Figs. 3.6 and 3.7 for two cases where $\sigma = \sqrt{10}$ and $\sigma = \sqrt{70}$. The primary peak at $\Delta/\Delta_r = 1$ corresponds to the correct step-size $\Delta = \Delta_r = 8$. As mentioned, the value of f_{ex} at the "true" peak should be close to zero. We define a "confidence interval" equal to σ_{ref} around the line $f_{ex} = 0$. A step size Δ is considered as a possible candidate when the value of $f_{ex}(\Delta)$ is closer than $\pm\sigma_{ref}$ to the line $f_{ex} = 0$. In Figs. 3.6 and 3.7 the area within the confidence interval is indicated by the shaded area. Fig. 3.6 shows that the primary peak $\Delta/\Delta_r = 1$ locates in the confidence interval while the secondary peak at the half step-size is characterized by the significantly larger value of f_{ex}.

Joint probabilities

When the signal-to-noise ratio is poor, the free energy curves may contain several peaks that approach the reference free-energy curve to within the statistical error. Even then it may still be possible to decide which peak is most likely to correspond to the underlying step size. The reason is that we consider a situation where there is a single underlying step-size. Hence, if one peak is "real" then the others are necessarily "over tones" or they are due to statistical fluctuations. Let us consider a specific case where we observe that the "measured" free energy is less than, say, two standard deviations removed from the reference free energy for step-sizes Δ_a and Δ_b. If the true step-size is equal to Δ_a, then we know the reference free energy at Δ_a (denoted by $f_a(\Delta_a)$) and its variance. But we can also compute a different reference free energy, namely the average free energy at step-size Δ_b given that the

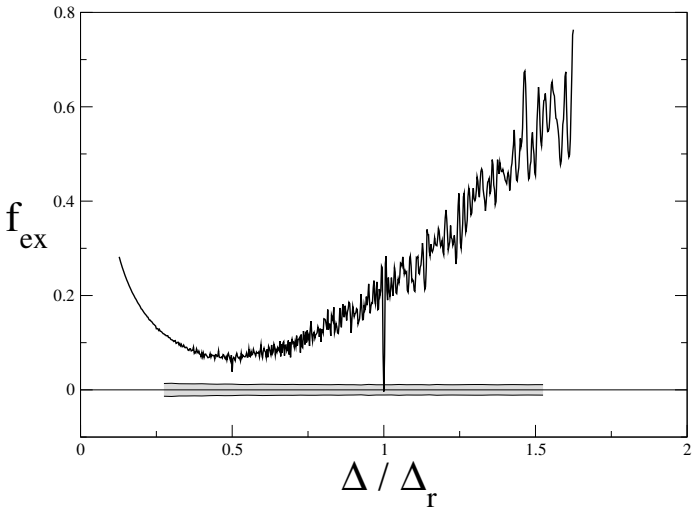


Figure 3.6: Excess free energy $f_{ex} = f(\Delta, 0) - \bar{f}_{ref}(\Delta, 0)$ as a function of the dimensionless step-size Δ/Δ_r for $\sigma/\Delta_r = \sqrt{10}/8$. The shaded area is the area of confidence interval defined as $\pm\sigma_{ref}$ around line $f_{ex} = 0$.

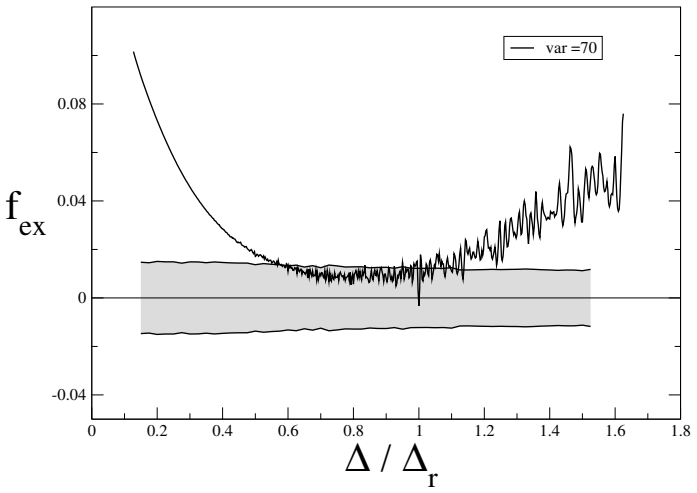


Figure 3.7: Excess free energy $f_{ex} = f(\Delta, 0) - \bar{f}_{ref}(\Delta, 0)$ as a function of the dimensionless step-size Δ/Δ_r for $\sigma/\Delta_r = \sqrt{70}/8$. The shaded area is the area of confidence interval defined as $\pm\sigma_{ref}$ around line $f_{ex} = 0$.

true step-size is Δ_a (denoted by $f_a(\Delta_b)$) and, as before, we can compute the variance of this quantity. Conversely, we can compute $f_b(\Delta_a)$, the average free energy at step-size Δ_a given that the true step-size is Δ_b, and its variance. In general, the statistical fluctuation in $f_i(\Delta_j)$ and $f_i(\Delta_k)$ will be correlated and we should also compute the cross correlations to construct an error ellipse. However, for the sake of simplicity we assume that fluctuations in the reference free energy are Gaussian and uncorrelated. It is then easy to compute the combine probability of observing the measured free energies at Δ_a and Δ_b, both for the case where Δ_a is the true step-size (we denote this joint probability by $P_a(\Delta_a) \times P_a(\Delta_b)$) and for the case where the real step-size is Δ_b (corresponding joint probability $P_b(\Delta_a) \times P_b(\Delta_b)$). The ratio $R = \{P_a(\Delta_a) \times P_a(\Delta_b)\}/\{P_b(\Delta_a) \times P_b(\Delta_b)\}$ then determines which of the two scenarios is more likely. Obviously, this procedure can be generalized to a larger number of peaks, but it is unlikely that this will be very useful. The most likely two-peak scenario is the one where we have a peak at Δ and at $\Delta/2$. The method described above should be able to deal with that situation. The generalization to correlated noise is straightforward. Later in this chapter we test the sensitivity of our method using the joint probability ratio R.

3.5 Sensitivity

Overview of existing methods

Let us first briefly mention some of the existing step detection methods. These are the "t-test" that has been described by Carter et al. [29], "velocity thresholding" as described by Hua et al. [31], Wavelet transform multi-scale products [36, 37], and χ^2 minimization [30]. A recent study by Kalafut and Visscher [38] also based on χ^2 minimization showed further improvement in the performance, and there are other works by Milescu et al. [32] that assumes a Markovian model for the stepping.

In a recent review Carter et al. [39] have compared the performance of the first four methods [29, 31, 36, 30]. Below, we briefly sketch the approach used in these most widely-used methods:

The T-Test method compares N points after and N points before each data point,

by introducing a t parameter:

$$t = \frac{m_{\text{after}} - m_{\text{before}}}{\sqrt{\frac{\sigma^2_{\text{before}}}{N} + \frac{\sigma^2_{\text{after}}}{N}}} \tag{3.15}$$

Where M is the mean and σ is the variance of each set. It is clear from the above equation that t will be the maximum when an upward step occurs and is minimum for a backward step. Steps are defined when the absolute value of t passes a threshold value.

Velocity Thresholding transforms the motor trace to the velocity profile by looking at different time frames. Steps are then defined when a velocity passes a certain threshold.

The χ^2 **minimization algorithm** gives the trace that has the lowest residual χ^2 by fitting the traces with increasing number of steps.

Some of these methods need one or more parameters from the experimental results to be able to correctly reproduce the trace. Different filters can be used to increase the performance of these methods, where a filter replaces the value of each point with the median or mean value of a certain window around that point. In their study Carter et al. [39] have concluded that the χ^2 minimization [30] method has a better performance than others at high level of noise. For this reason we compare the technique that we propose with χ^2 minimization. Here we describe the χ^2 minimization method in more details and in the next subsection we compare our free energy method to this method.

The χ^2 minimization algorithm starts by fitting a trace with only one step to the experimental data. The best fit is the one with the lowest χ^2 and gives the position and size of a step which most likely locates at one of the true steps of the underlying data. Next the algorithm does exactly the same on each resulting plateau on the left side and right side of the first step and finds the locations of new steps. This is continued until only few data points are left for each steps. Fig. 3.8 (a)[30] shows the fitting procedure. While in the beginning the traces under-fit the data, they start to over-fit the data as the number of steps increases in the last iterations.

In order to refine the algorithm and evaluate the quality of each of the fitted traces Ref. [30] introduces a so called S value. This value measures the difference between the quality of a fit and its "counter-fit". In order to get a counter-fit the same best fit algorithm is performed with a difference that at each iteration after finding the

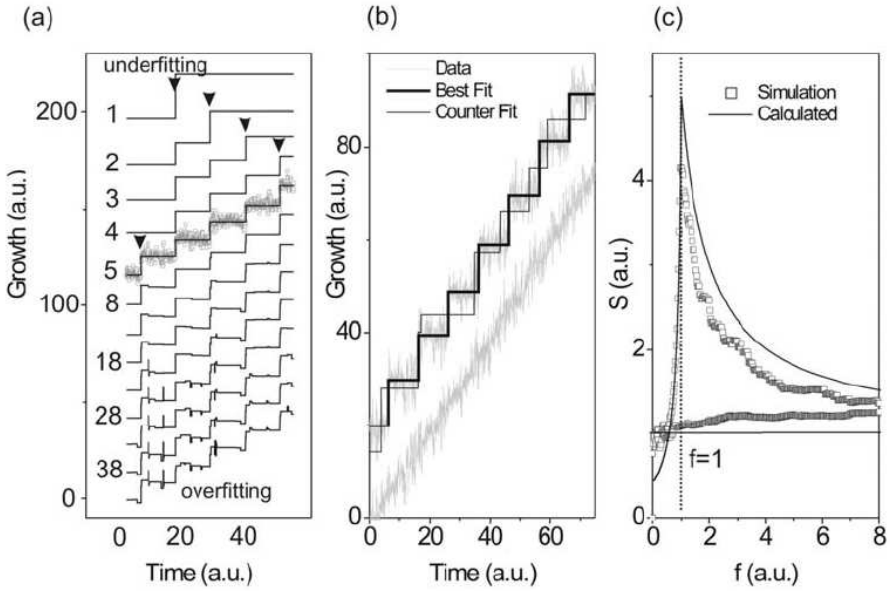


Figure 3.8: Step fitting procedure from Ref. [30]. a) Iterations of the step fitting algorithm. The arrowheads show every new step that is added in each iteration. b) A best-fit together with a counter-fit for a stepping data [30]. The lower curve shows linear non-stepping data c) S parameter versus the relative number of fitted steps from simulation results of Ref. [30]

position of the new step, the original step locations on the plateau will be rejected. The outcome is a trace in which all the step locations are misplaced. The quality of this counter-fit and the best-fit thus differ strongly. The method uses this to measure the quality of fits. The S parameter is then defined as the ratio of χ^2 of the counter-fit to the best-fit.

$$S = \frac{\chi^2_{\mathrm{counter-fit}}}{\chi^2_{\mathrm{best-fit}}} \tag{3.16}$$

If there is severe over-fitting or under-fitting, or if the original data consists of non-stepped trace, then S will be closer to 1. S will be maximum when the quality of fit is the best. In their study, Kerssemaker et al. show that the value of the peak is roughly given by $1+\Delta^2/(4\sigma^2)$ when the fit is optimum. Carter et al. [39] show that compared to other algorithms the χ^2 minimization has the best performance especially when the signal-to-noise level (Δ_r/σ) approaches 1.6, the lowest level tested.

Sensitivity of free energy method

We have tested the sensitivity of our method on several sets of artificial data of $N = 4000$ data points with different levels of noise. The underlying URWs were generated from a Poisson distribution with the average time $\tau = 100$ between each two successive steps, and a fixed step-size $\Delta_r = 8$. We expect that at a sufficiently high level of noise, the peak of the excess free energy will submerge in the noise of the free energy profile.

For comparison, we applied the χ^2 minimization algorithm to two of our artificial data with level of noise $\Delta/\sigma = 2.52$ and $\Delta/\sigma = 0.96$. Figs. 3.9 and 3.10 show $S - 1$ as a function of step-size for these two data sets. The arrow in the figures show the position of the correct step-size. As is clear from the figures irrelevant peaks already start to compete with the correct peak in the case of $\Delta/\sigma = 2.52$. These peaks become totally dominant at the larger level of noise $\Delta/\sigma = 0.96$ as is shown in Fig. 3.10. To compare to our algorithm we refer to Figs. 3.6 and 3.7 where we applied our free energy method to the very same artificial data sets that had been used in Figs. 3.9 and 3.10. These results suggest that the free energy method is more efficient than the method of Kerssemakers et al. which works well for the signal-to-noise ratio down to $\Delta/\sigma \approx 1.6$.

In order to test the sensitivity of the algorithm we calculate the ratio R of the joint probabilities for the most likely two-peak scenario where we have two peaks

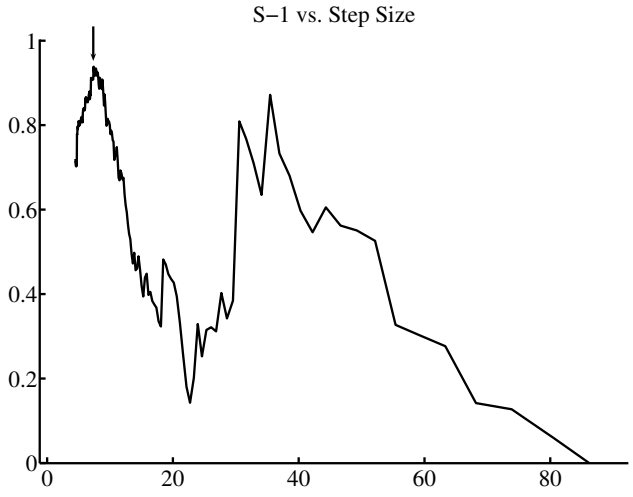


Figure 3.9: s-1 versus step-size for $\sigma = \sqrt{10}$ and $\Delta_r = 8$ ($\Delta/\sigma = 2.52$) from χ^2 minimization method, the peak shows the best fit and the arrow shows the correct value for the step-size. The correct peak already competes with other peaks.

Figure 3.10: s-1 versus step-size for $\sigma = \sqrt{70}$ and $\Delta_r = 8$ ($\Delta/\sigma = 0.96$) from χ^2 minimization method, the peak shows the best fit and the arrow shows the correct value for the step-size. The peak at the correct step-size is not dominant.

Table 3.1: $R(\Delta_r, 0.5\Delta_r)$, the ratio between the likelihood of step-size Δ_r to likelihood of step-size $\Delta_r/2$ for several levels of noise.

Δ/σ	$R(\Delta_r, 0.5\Delta_r)$
2.52	10^{81}
1.78	2.82×10^7
1.14	5.09
0.96	1.99
0.68	1.10

at Δ_r and at $\Delta_r/2$. Table 3.1 shows the $R(\Delta_r, 0.5\Delta_r)$ which is the ratio between the likelihood of step-size Δ_r to the likelihood of step-size $\Delta_r/2$ for several levels of noise. Since we "know" the true step-size is Δ_r we expect R to be larger than 1 when the noise level is sufficiently low. The results shown in table 3.1 suggest that the joint probability argument correctly predicts the true step-size even at a very high level of noise $\Delta/\sigma \approx 1.0$.

To Summarize, in this chapter we proposed an algorithm to estimate the most likely step-size compatible with noisy experimental data on the motion of molecular motors along linear "tracks". Our method is based on the numerical estimate of the "free energy" of an ensemble of directed random walks of different step sizes. We show that the method is capable of detecting a single step size at a signal-to-noise ratio that is significantly lower than the ones that can be studied using the χ^2 minimization [30] method. In the simple version presented in this thesis, the method that we use is limited to processes with a single underlying step size and sign. In contrast, the χ^2 minimization [30] method can be used to analyze multiple step-sizes traces or traces with forward or backward steps. This limitation of the present method is not inherent in the approach. As we briefly discussed in section 3.3, our approach can be generalized to more complex situations. The performance of all methods to analyze stepping motor traces depends on the size of the experimental data set. All the results that are reported in this chapter are based on data sets that contain $N = 4000$ data points with, on average one hundred data points between successive steps. It should be stressed that the comparison between the present approach and other methods may, in principle, depend on the size of the data set and on the number of points per step.

Acknowledgment This chapter is based on the collaborative work between Kostya Shundyak, Myself, and Daan Frenkel. I would like to thank Kostya and Daan for

many daily disscussions.

4 A Coarse-grained model for DNA coated colloids

4.1 Introduction

The highly specific nature of complementary-strand binding in double-stranded (ds) DNA offers many possibilities for the use of this biomolecule in micro- and nano-technology. For example, colloidal particles coated with distinct species of short (around 100 bases) single-stranded (ss) DNA (called spacers) can bind specifically with each other via suitable DNA 'linkers' [7, 8, 9, 10]. These "suitable" linkers are themselves ssDNA strands added to the solution. Such DNA-coated colloids may offer a route to novel colloidal crystals [8, 10]. To make colloidal particles crystallize is an interesting problem in its general form, since it is one of the promising ways to control the structure and therefore properties of materials. This is not a new phenomenon for experimentalist: Without using any coating spacers or linkers, many particles do make ordered structure (as an example when the surrounding solution evaporates). However, these self-Assembly of particles is very sensitive to their chemical property of particles. The polydispersity in shape of the particles, charge, and chemical structure of particles together with short range van der Waals interactions rule the properties of the assembly. Practically these "unspecific" and uncontrollable interactions is one of the main challenges in controlling the structure of self-assemblies. For this reason, the very "specific" nature of binding interactions in DNA linkers becomes as a key. The four basic building blocks of DNA, i.e. the bases adenine (A), cytosine (C), guanine (G), and thymine (T), placed along the backbone of ssDNA strand, have a very specific interaction among themselves. According to Watson-Crick model [40] the A groups can only bind to T groups while the C groups can only bind to G. These base-pairing interactions makes a ssDNA strand bind to its "complementary" strand and form a double-stranded DNA. Therefore, careful designing of the sequences along the DNA strands, (in which one can safely ignore partial binding and mismatches) can introduce a one-to-one interaction between "a" ssDNA strand and its "only" complementary strand.

In experiments it is possible to attach these ssDNA strands to colloidal particles

(See section 4.2) and make them "sticky". These sticky parts will interact with each other in the same way that isolated ssDNA strands interact. Therefore ideally one can introduce the same level of specificity between colloidal particles that are coated with complementary strands of ssDNA. Importantly, the specific DNA-mediated interactions may be made stronger than the non-specific interactions that normally dominate aggregation and self-assembly in colloidal systems. Recent experimental studies on gold nano-particles [5, 6] have successfully used short DNA strands to "direct" the self-assembly of particles into a non-close packed crystalline structure. Nevertheless, experiments that aim to achieve DNA-controlled self-assembly are very challenging and our understanding of the factors that control DNA-mediated interactions between colloids is still far from complete. One factor that plays a crucial, but incompletely understood, role is the length of the DNA polymers that have been grafted onto the individual colloids. Recent experiments on colloids coated with relatively long DNA strands [41, 42] show that, depending on the length of the DNA, the colloids may either assemble into finite, disordered clusters or into system spanning aggregates, Other factors that play a key role in DNA-mediated self-assembly are the temperature [5, 6], the grafting density, and the ionic strength of the solvent [43]. Of course, for experiments, the size of the colloid itself is an important variable because direct optical imaging of large (micron-sized) colloids is straightforward, whereas nano-colloidal assemblies must be probed either by X-ray scattering or by electron microscopy. In what follows, we shall focus on the behavior of micron-sized colloids coated with DNA. To date, most of the experiments on such systems of large, DNA-coated colloids have considered colloids coated with relatively short DNA - usually with a contour length that was less than the persistence length of the DNA. Under those circumstances, DNA behaves like a rod, rather than a polymer coil. However, the experiments of refs. [41, 42] focused on systems of colloids coated with DNA's that contained many thousands of base pairs and that behaved as a polymer coil with a radius of gyration that was comparable to the size of the colloid. One should expect that the use of long and very extensible DNA linkers between colloids will qualitatively change the way in which the colloids self-assemble. From a theoretical perspective, the phase behavior of colloids coated with DNA coils was considered by Tkachenko [44] who found that the combination of screened Coulomb repulsion between the colloids and reversible, DNA-mediated attraction between particles coated with complementary DNA strands could lead to the formation of various non-close-packed crystalline phases, such as for instance a diamond structure. The confocal-microscopy studies of Schmatko et al. [42] on colloids coated with DNA that had a radius of gyration similar to the colloid radius, showed that these colloids

formed aggregates with considerable local compositional order (i.e. a given colloid tended to be surrounded by colloids with complementary DNA coating) but no overall crystalline structure. was observed.

Here we present a simulation study of a model inspired by the system of colloids coated with long complementary DNA's studied in Ref. [42]. As the colloids studied in the experiments are micron sized and contain of the order of 10^{10} atoms, whilst every DNA strand contains many thousands of base pairs, it is obvious that a fully atomistic simulation of the phase behavior is utterly unfeasible with existing (or even foreseeable) computing resources. We therefore must construct a "coarse-grained" model that can be studied by Monte Carlo simulation. In constructing this model, we specifically use the fact that DNA molecules are very long. During the past decade, Hansen, Louis, Bolhuis and Meijer have analyzed the problem of the description of effective (coarse-grained) interactions between long polymer chains in a good solvent (see e.g. Ref. [24]). A key result of this work is that such polymer chains behave effectively as soft spheres. The strength of the effective interaction between these soft spheres depends on concentration but in the low density regime that we shall focus on, it is well described by a Gaussian repulsion with a strength of approximately $2k_BT$. Replacing the polymers in a molecular solvent by soft, repulsive spheres in a uniform background constitutes a huge simplification of the model for the polymers. Similarly, we will simplify the description of the colloids by assuming that they behave as hard spheres. This approximation is justified for uncharged (or strongly screened) colloids in a medium that matches the refractive index of the colloid. In the experiments, the second condition may not be perfectly satisfied and this will result in certain differences between simulation and experiment to which we will return later. In the remainder of the present chapter we first discuss the experiment of Schmatko et al.[42]. Then we present our coarse-grained model and we discuss the simulation techniques used in this study. Novel predictions about the phase behavior of coated-colloids will be presented in the next chapter.

4.2 Long DNA experiments

Schmatko et al. [42] reported an experimental study of the aggregation of colloidal particles that could be linked by hybridization of long strands of grafted dsDNA that was functionalized by short, complementary ssDNA "sticky ends". In this experimental work, the authors studied the aggregation of 1μm colloids bridged by DNA with 32μm contour length, where the "contour length" is defined as the total length of DNA strand along its backbone. In the experiments, two species of col-

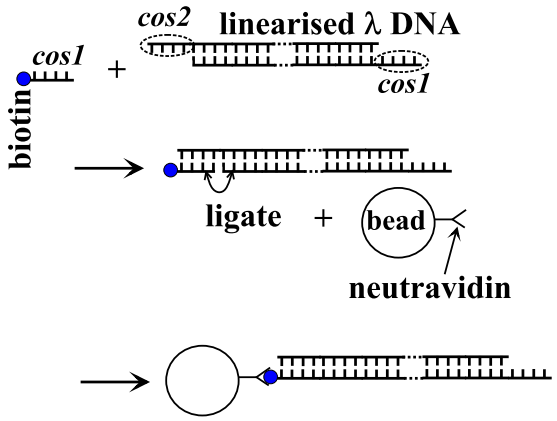


Figure 4.1: Schematic representation of colloid particle coated with ds λ-DNA through Biotin-neutravidin connection.

loids with grafted double-stranded λ-DNA displaying short, complementary single-stranded 'overhangs' as free binding-ends, were mixed. The λ-DNA is a long ds-DNA which is predominantly circular at room temperature and has a contour length of 16 μm (roughly 48,500 base pairs (bps)). Upon heating to 65°C, the ds-DNA breaks open and linearizes. Each of these linear ds-DNA is now terminated by two complementary 12-base single strands that we call "sticky ends". A schematic summary of the particle-preparation protocol is given in Fig. 4.1. In the experiments, two populations of linear DNA were prepared. In each population, one sticky end was hybridized and ligated to biotin [1]. These two populations of DNA were then separately mixed with neutravidin-coated green-fluorescent ('A' hereafter) and red-fluorescent ('B' hereafter) polystyrene colloids (diameter 1 μm). Using dynamic light scattering the authors of Ref. [42] estimated the radius of the linear λ-DNA coil to be roughly 800 nm .They also estimated that some 10 λ-DNA chains were grafted onto each colloidal particle. Given the relative sizes of the DNA and particles, one would then expect each colloidal particle to be surrounded by a "cloud" of grafted DNA. Subsequently, equal amounts of the complementary colloids were mixed, reaching a final particle volume fraction of $\phi \approx 0.004$. The suspension was then loaded and sealed into a sample chamber. After several hours, during which the DNA strands could start to hybridize, some clusters consisting of around 10 particles had formed, mixed with a significant number of smaller clusters. Summarized, the experiments of Schmatko

[1] *Biotin is a water-solvable vitamin that has a high affinity for proteins named avidins, such as neutravidin.*

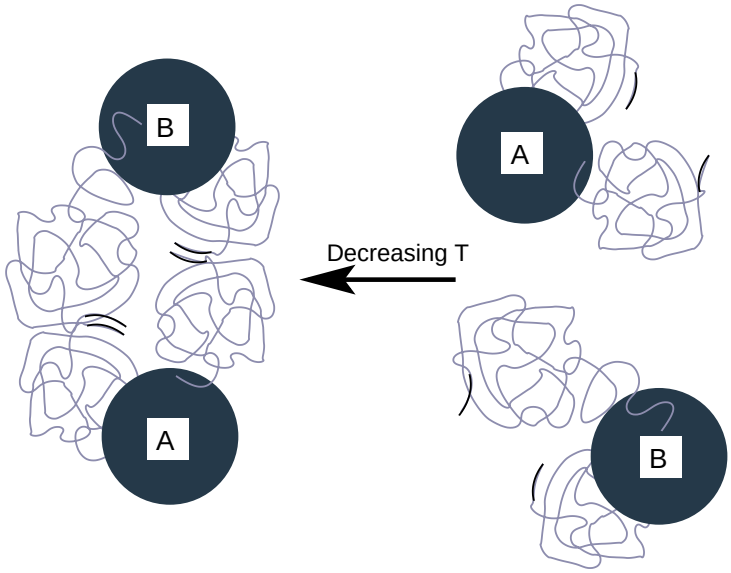


Figure 4.2: Schematic representation of the hybridization reaction of two types of particles (labeled as A and B) coated with long dsDNA chains. The short dark-colored terminal parts of the chains are ssDNA threads which are complementary for A and B type of colloids. Upon decreasing temperature the terminal parts of DNA strands hybridize and introduce a physical attraction between colloids of different type. Clustering of A-B particles have been observed in experiments because of this hybridization of complementary end groups[42].

et al. [42] report the formation of stable, size-limited clusters of DNA-coated colloid. These clusters have two main characteristics: One is that they fail to grow further in size: they remain size-limited even after many days. Second, inside each cluster the two species are surprisingly close together - in fact, many particle pairs appear to be touching. This means although the DNA strands are very large (their size is comparable with colloids), the two connected colloids stay very close together. This suggests that the DNA strands get excluded from the space between colloids. More discussions on the results and supporting theory can be found in Refs. [42, 45].

In a more recent experiment, Geerts et al.[41] studied the effects of the length of dsDNA tethers on clustering of colloids. They considered three different cases, i.e. very long (48,500 base pairs (bps)), intermediate but still long (7500 bps), and no dsDNA. In the last case, the complementary ssDNA parts are directly grafted on to the colloidal particles. They found system-spanning clusters of coated colloids even

in the intermediate regime.

4.3 The model

The coarse-grained model that is used in our simulations is meant to capture the main characteristics of the experimental system described in the previous section. In general, the system can be categorized as a mixture of colloids and grafted polymers with "sticky" ends. Here, we list some of the important characteristics of the system:

- - **The system consists of two types of colloids**. Labelled as A and B (equivalent to the green and red fluorescent labels in the experiment). Colloids of different species attract each other after hybridization occurs between their complementary polymer tethers. Colloids of the same kind only experience repulsion. It is an important characteristic of the system that differences in interactions between the colloids are only due to the difference in DNA coating and not due to differences in size or shape.

- - **The DNA-strands are long**. As we will discuss in detail in the following section, we model this DNA strand as "soft" particle with an effective size R_g, where R_g is the radius of gyration of the DNA molecule. We therefore, characterize a DNA chain only by the position of its center of mass.

- - **The hybridization reaction is specific and reversible**, and its degree of reversibility depends on temperature. This comes from the fact that the sticky ends of polymers molecules can bind and unbind only to their complementary partners.

The model that we use takes the characteristics listed above into account. We consider a binary mixture of N_c hard spheres (colloids) of type A or B, each covered with n identical DNA strands. In this model polymers are permanently tethered to the colloids and there are no free polymers in the system . Therefore, the total number of DNA strands is $N_p = nN_c$. For the sake of simplicity, we assume at this stage, that the number of DNA's per colloid is always exactly n. In experiments, it is likely that the number will fluctuate. Although it would be straightforward to generalize the model to take fluctuations in n into account, we start here with the simplest case, viz. fixed n.

For the Monte Carlo simulations we introduce two types of moves: first of all, we perform normal Monte Carlo trial moves that attempt change the center-or-mass position of either the colloids or the polymers. In addition, we implement reaction moves

that simulate the hybridization and dehybridization of the complementary DNA's that are grafted to different colloids. The schematic representation of a hybridization reaction is described in Fig. 4.2, and a cartoon of soft sphere picture representation is presented in Fig. 4.3.

Now that we have specified the components in our coarse-grained model, we still have to specify their interactions. One part is straightforward: we model the interactions between the bare colloids assuming that these behave as hard spheres. The interaction between the polymers is more subtle: the soft-colloid picture for polymers in a good solvent [24] was developed for polymers that are not tethered to a colloid, nor for the situation where the polymers form a bridge between two colloids. For the present, purely exploratory model we make the simplest possible assumption, namely that the interaction between tethered polymers is the same as that between free polymers. In addition, we assume that we can always use the low-density limit for the effective polymer-polymer interaction potential. Finally, we assume that, although the colloids repel the polymers (see below), they do not change the interaction between the polymers. These assumptions are the simplest ones that one could make. They are almost certainly an oversimplification. We speculate that a more quantitative description of the polymer-polymer interactions will affect the quantitative, but not the qualitative aspects of the behavior that we observe in our simulations. Whether this speculation is justified remains to be seen.

Of course, the "soft-colloid" picture for polymers does not account for the hybridization of sticky ends. In our model, we will assume that the sticky end of a polymer is bound harmonically to the center of mass of the polymer, but does not otherwise affect the properties of that polymer. Similarly, we will assume that the end of the polymer that is grafted to the colloid is also harmonically bound to the center of mass.

In the rest of this section we discuss in details the "effective" pair-potentials and the Monte Carlo reaction moves for tethered polymers in their "center of mass" representation.

4.4 Pair-potentials for grafted, sticky polymers

The soft picture model for polymer chains has been already discussed in literature [24, 46]. Bolhuis et al.[24] studied the effective interactions in dilute or semi-dilute solutions of polymer chains. As a part of their work, they proposed effective pair-potentials between centers of mass of two polymer chains, as well as an effective form of the interaction between a polymer chain and a flat, hard wall. Pierleoni et

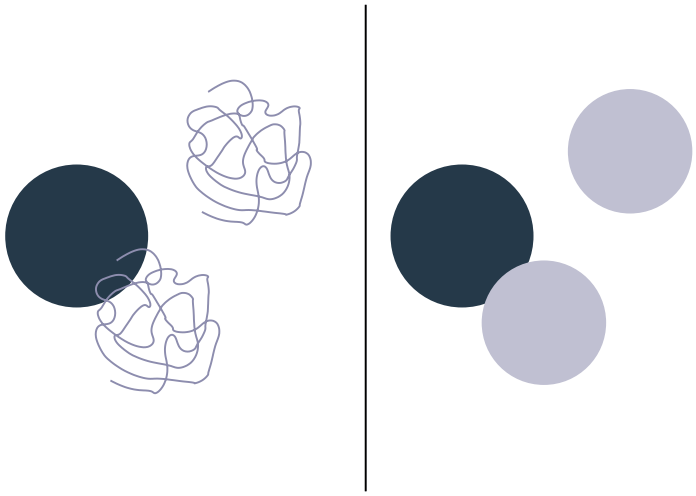


Figure 4.3: Soft picture representation for long polymer coil. Dark-color spheres represent hard colloids and light-color ones are the soft penetrable spheres representing polymer chains.

al. [46] studied pair-potentials for block copolymers in the blob picture. In the blob picture [15] a strand of long polymer is represented as a string of blobs. Each blob then represents part of the chain. Pierleoni et al. calculated the effective interactions for tethered and untethered blobs in such a chain. Both studies are relevant to the problem of DNA-coated colloids: One studies the pair-potential for free polymers, while the other studies the effect of tethering. The approach that we take here is more simplistic than the multi-blob picture proposed in Ref. [46]: we approximate a grafted polymer as a single blob. Again, it is obvious that a more sophisticated description is possible (at a cost) and would certainly be called for in situations where the grafting density is high. However, as the present model is meant to be a minimal model, we limit ourselves to the simplest possible description.

Consider the zero density limit of DNA-coated colloid system, in which we have only two hard colloids each coated with a single DNA strand with complementary ends. To stay close to the experimental model presented in the previous section let us assume that the DNA strands are sufficiently long so that we can safely use the soft picture representation. Now the following interactions have to be considered:

1. interaction between two hard colloids
2. interaction between two polymers

3. interaction between a polymer and a colloid

As mentioned above, we assume the bare colloid-colloid interaction V_{c-c} to be a hard core interaction that suppresses any overlaps between colloids. This will be the same between all colloids and does not depend on their species:

$$V_{c-c}(r) = \begin{cases} \infty & \text{if } r \leq 2r_{col} \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

where r is center-center distance and r_{col} is the radius of colloids. A and B colloids are assumed to be identical in size.

Within the soft-colloid picture [24], the pair-potential V_{p-p} between the centers of mass of two free polymers in a good solvent can be described to a very good approximation by a Gaussian model given by:

$$V_{p-p}(r) \approx 2k_{\mathrm{B}}T \exp\left[-0.7(\frac{r}{R_g})^2\right]. \tag{4.2}$$

In their study, Bolhuis et al. show that V_{p-p} is transferable and, as long as the concentration of the polymers is well below the overlap concentration, V_{p-p} can be assumed to be constant. However, in our model DNA strands are permanently connected to the colloidal spheres with one end and they may bind to another DNA strand with the other end when hybridization occurs.

This tethering between particles modifies the pair-interactions. First, let us look at the effect of hybridization: When two DNA strand hybridize they lose some degree of freedom because of the constraint on their end points. This loss in entropy introduces an attractive pair-interaction between two polymers. For two ideal chains this interaction takes the form of an harmonic oscillator which is a quadratic function of the end to end separation.

It is, of course, not a priori obvious that one can assume the additivity of the short-ranged repulsion between two polymers and the entropic attraction due to bond formation. However, support for this conjecture comes from the work of Ref. [46] that shows that such additivity is reasonably well satisfied in a close related model.

In what follows, we therefore assume the following expression for the coarse-grained interaction between the polymeric chains:

- The pair-interaction between two polymers that are not connected to each other is assumed to be of the form $V_{p-p}(r)$ (eqn. 4.2).

- The pair-interaction between two polymers that are connected end-to-end is assumed to be of the form $V_{p-p}(r)+V_{ho}(r)$, where V_{ho} has the form of harmonic oscillator.

Let us now look at the third category of interactions: Hard spheres and polymers. In general, in the studies of colloid-polymer mixture, two limits of size ratios are mostly discussed. One is the so-called colloid limit which includes large colloids and small polymers with a radius of gyration much smaller than radius of the colloid. In this limit the pair-interaction between a polymer and a colloid can be related to the interaction between a polymer and a wall using the Derjaguin approximation [47]. The opposite limit, which is sometimes referred to as the protein limit, is obtained when the colloids are much smaller than polymer chains. In an extreme limit of this case the interaction between the colloid and the polymer should vanish as the radius of the colloid goes to zero. The most interesting case for us, however, is the intermediate regime in which the radius of polymer and the colloid are comparable. Bolhuis et al. [24] present effective pair-potential interactions of a free polymer chain close to a hard sphere for one size ratio in the intermediate regime. Pelissetto and Hansen [48] carried out lattice Monte Carlo simulation for a wide range of size ratios. In part of their studies they present the effective pair-potential in the infinite-dilution limit in an athermal colloid polymer mixture. They determine the effective two component pair-potential between centers of mass of an isolated polymer on lattice and a hard colloid in different limits where $R_C/R_g = 0.02-10$. However, unlike the polymer-polymer interactions which has effectively a Gaussian form, polymer-colloid interactions have not similarly been expressed in a closed form. Even for the case of ideal polymers, the pair-potentials are not known analytically in the center of mass representation. Moreover, "tethering" also plays a role here: In our model, DNA strands are permanently connected to the surface of hard colloidal spheres with one of their end groups. This will introduce an attractive harmonic spring interaction between the colloid and the polymer and, as we indicated above, the additivity of the short range repulsive interaction and the long range attractive part need not hold.

The aim of the present study is not to perform an exhaustive study of the interactions between colloids and tethered polymers as a function of size ratio. For the present work, a workable description of the polymer colloid interaction suffices. In order to achieve such a description for the polymer-colloid systems that we study, we have to collect numerical data that allow us to construct an effective pair potential. To this end we have performed Monte Carlo simulations similar to those of Pelissetto [48] of tethered polymer-colloid pairs at infinite dilution.

4.4.1 Colloid - polymer interaction at zero density

The first step towards the construction of transferable polymer-colloid potentials is to compute the interaction between a hard colloid and a free polymer chain. From these simulations we can calculate the zero-density limit of the potential of mean force that acts between the polymer and the colloid. Next, we report results for single polymers tethered to the surface of the sphere with one end. We compare these two pair-potentials to check the additivity, and the effects of tethering.

In the zero density limit the potential of mean force that we derive from the polymer-colloid radial distribution function $g(r)$, is equal to the effective pair potential $U(r)$.

$$\frac{U(r)}{k_{\mathrm{B}}T} = -\ln g(r). \tag{4.3}$$

Let us first consider the end-to-end distribution of ideal chains. In the regime where the end-to- end distance is much smaller than the full extension of the ideal chain the Gaussian approximation is valid for the end-end distribution [49]. This will immediately give the quadratic spring form for the free energy of the chain at fixed elongation

$$U(r) = -\frac{3k_{\mathrm{B}}T}{2N}\left(\frac{r}{b}\right)^2, \tag{4.4}$$

where N is the number of monomers in the chain and b is size of the monomer. For an ideal chain the mean-square radius of gyration is given by:

$$\langle R_g^2 \rangle = \frac{Nb^2}{6}, \tag{4.5}$$

and, therefore we can rewrite $U(r)$ as

$$U(r) = -\frac{k_{\mathrm{B}}T}{4}\left(\frac{r}{R_g}\right)^2. \tag{4.6}$$

If we express all distances in units R_g, the coefficient $\kappa = 0.5k_{\mathrm{B}}T$ is the "entropic spring constant" of an ideal chain.

As can be seen from Eqn. 4.6 the spring constant of an ideal linear spring is of the order of $k_{\mathrm{B}}T$. For other situations than the end-to-end distance of an ideal chain (e.g. in the case of the center-of-mass distribution of two end-connected polymers) we expect that the interaction potential is still to a good approximation harmonic. However, the spring constant will in general be different from the one given in eqn. 4.4. In other words, the value of the spring constant is one of the control parameters of the

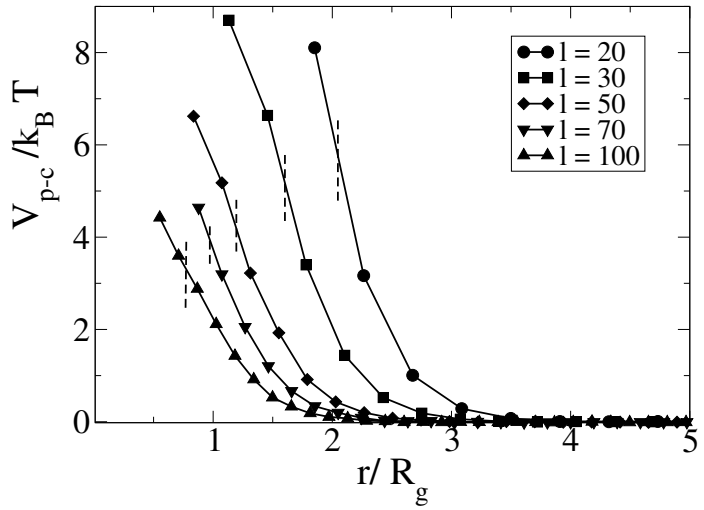


Figure 4.4: Effective Colloid-polymer pair-potential from lattice simulations for a free polymer next to a sphere with radius r_c5. Different curves correspond to different length l of the polymer. Radius of Gyration of polymers are between $R_g/r_c = 0.49 - 1.27$ for $l = 20 - 100$. The dashed lines indicate the point where the polymer center of mass penetrates the hard colloid. Fitting the curves to a single Gaussian we find out that the value at full overlap decreases as the length of the polymer increases.

system. It affects the strength of the attractive interaction between bound polymers. For linear polymers, the size of the polymer and the strength of the entropic spring constant are directly coupled. However, polymers with a different topology (e.g. star polymers) will have a different value of κ for the same value of R_g. In the remainder of this chapter we chose to work with linear chains.

To compute the effective interaction between a linear polymer (bound and unbound) and a hard colloid, we carried out Monte Carlo simulations for linear chains of length between $l = 20$ up to $l = 100$ lattice spacings, using Configurational Bias scheme on lattice [11]. The length of the box was chosen to be $L_{box} = 100$ and periodic boundary conditions are applied to all the directions. In this lattice representation, the sphere is defined as all the points that are closer to the center than the radius of the colloid, $r_c = 5$, lattice spacings. These points are inaccessible to the monomers along the polymer chain. We sample a configuration every 1000 Monte Carlo moves and calculate the distribution function of the polymer center of mass as

well as the radius of gyration. In the same way as Eqn. 4.3 the pair potential between the sphere and the polymer is related to center-center distribution function of colloid and polymer. We study two cases: a free polymer next to a sphere and a polymer that is tethered with one end point to the surface of the sphere. In this case, we fix the first monomeric unit of the polymer (i.e the end that is tethered) on one of the empty lattice sites in the first layer around the hard sphere.

Fig. 4.4 shows the computed pair-potentials that act between a hard colloid and an unbound polymer, as a function of the radius of gyration of the polymer (at constant colloid size). The minimum size ratio is $R_g/r_c = 0.49$ for a polymer of contour length 20 and the maximum size ratio is $R_g/r_c = 1.27$ for a polymer of length 100. The distances in each profile are scaled with the radius of gyration of the polymer in solution. The results of this simulation show that even for the smallest size ratio $R_g/r_c = 0.49$ the center of mass of the polymer can penetrate the colloid to some extent (something that is clearly impossible for a polymer approach a hard wall). In the figure, the position of the colloid surface is indicated for each l. As the length of polymer increases the value of the potential at surface decreases. This is not surprising because long polymers can wrap arround the colloid. Our results are qualitatively in agreement with Pelissetto's [48], however, in our simulation polymer's length are much shorter than the extensive work of Pelissetto. Next, we consider a polymer chain of length $l = 100$, which is tethered to a colloid with $r_c = 5$ via one end group. As Fig. 4.5 shows, the pair-potential V_{p-c} in this case is repulsive at short distances and has a quadratic attractive spring-like form at longer separations. This figure allows us to test to what extent the polymer colloid potential can be written as a sum of a harmonic attraction and a Gaussian repulsion (Eqn. 4.2). For comparison, Fig. 4.5 also shows the repulsive colloid-polymer pair potential for an unbound polymer. When we subtract the $V_{p-c}^{\rm free}$ for free polymer from $V_{p-c}^{\rm bound}$ for bound polymer, we obtain a function that, if the interactions are additive, should be a quadratic function of the distance between the polymer center and the colloid surface. We have verified this for two of size ratios $R_g/r_c = 1.27$ and $R_g/r_c = 0.838$, fitting $V_{p-c}^{\rm bound} - V_{p-c}^{\rm free}$ to a quadratic function. The results of this comparisons indicates that, at least for the cases that we studied, the pair-potentials are to a good approximation additive. We do not wish to imply that additivity will hold for all polymer-colloid size ratios (almost certainly, it does not). However, for the present study that focuses on polymer-colloid systems with size ratios similar to those used in the experiments of refs. [42, 41], additivity seems to be a reasonable assumption. We note that we cannot easily extract the value of the polymer-colloid pair potential at $r = 0$ from the data shown in Fig. 4.4. For the subsequent simulations, this point is of minor impor-

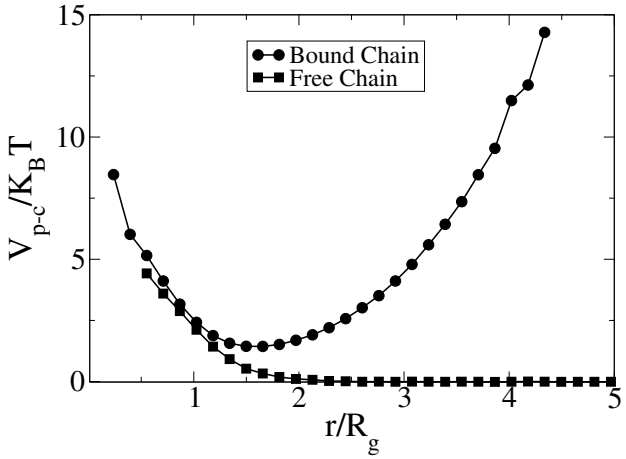


Figure 4.5: An example of a typical colloid-polymer pair-potential obtained by lattice simulations for a bound chain and a free chain. Distance r is the center-center distance; size ratio is given by $R_g/r_c = 1.27$ for a polymer of length $l = 100$.

tance, as the main point is that the interaction becomes strongly repulsive at small r. Configurations with r close to zero will therefore be rarely sampled. For the coarse-grained simulations reported in the next section, we fit the repulsive polymer-colloid potential to a Gaussian form

$$V_{cp}(r) \approx a \exp(-r^2/c).$$

4.5 Reaction moves

4.5.1 Hybridization equilibrium of DNA

Having discussed pair-potentials, we now need to develop a Monte-Carlo scheme that combines reaction moves, during which polymers bind or unbind, with displacement moves. Colloids of different types can be linked if their respective DNA strands hybridize. This can be described by a two state model for a dimerization reaction in a solution of short ssDNA segments A and B that can hybridize to form a short dsDNA strand AB:

$$A + B \rightleftharpoons AB \tag{4.7}$$

Later on, we will identify the two reactant "molecules" A and B as the two complementary single-stranded ends of a long DNA chain. The forward move represents hybridization leading to the product AB which is a double-stranded DNA chain. The backward move brings the system back to the random coil state of two single stranded chains. This two-state model is often used to estimate the temperature-concentration dependence of the dissociation of short, double-stranded DNA molecules into free ssDNA chains. It has been suggested [50] that this dimerization model provides a reasonable approximation of the melting temperature even when the duplex have a non-two-state transition, when DNA molecules form alternative conformations between duplex to random coil transition.

In the two-state model the condition for chemical equilibrium, the chemical potentials of the reacting species should satisfy

$$\mu_A + \mu_B = \mu_{AB} \tag{4.8}$$

If we assume that both the ssDNA and the dsNDA solutions are ideal we can relate the equilibrium constant for the hybridization reaction to the partition functions of reactants and products. For species x ($x = A$, B, or AB) the partition function is given by

$$Q_x = \frac{V^N}{N!} \frac{q_{int}^N(x)}{\Lambda_x^{3N}}, \tag{4.9}$$

where $q_{int}(x)$ is the contribution to the partition function of x due to all internal degrees of freedom. Note that the thermal De Broglie wavelength Λ_x depends on the mass of species x. The equilibrium condition can be rewritten as

$$\left(\frac{\rho_{AB}}{\rho_A \rho_B}\right) = \frac{q_{int}(AB)/\Lambda_{AB}^3}{\left(q_{int}(A) q_{int}(B)\right)/(\Lambda_A^3 \Lambda_B^3)}, \tag{4.10}$$

where ρ_x is the number density of species x. To compare with experiments, it is conventional to express the equilibrium constants such that the concentrations can be expressed in mol/liter:

$$\frac{[AB]}{[A][B]} = K, \tag{4.11}$$

where K is the equilibrium constant of the A-B dimerization reaction. If $[x]$ is the concentration in mol/liter, then $[x] = \rho_x/\rho_0$, where ρ_0 is the "standard" number density corresponding to one mol/liter, i.e. 6.022×10^{26} molecules per cubic meter. In this way we can relate the equilibrium constant to number densities via

$$\left(\frac{\rho_{AB}}{\rho_A \rho_B}\right) = \frac{K}{\rho_0}. \tag{4.12}$$

This step is necessary because, for a dimerization reaction, the equilibrium constant is not dimensionless. Using equation 4.10 we then obtain,

$$\frac{q_{int}(AB)/\Lambda_{AB}^3}{q_{int}(A)\ q_{int}(B)/(\Lambda_A^3\ \Lambda_B^3)} = \frac{K}{\rho_0} \tag{4.13}$$

Next, we will use the above relations to arrive at an expression for the probability of dimerization of to long chain molecules terminated by reactive A (B) ends. For the reaction algorithm of the Monte Carlo scheme we need the dimerization probability of two ssDNA chains. Consider the situation where A and B are the endpoints of DNA chains with centers of mass located at R_A and R_B respectively. We denote the center-to-center distance $R_A - R_B$ by R_{AB}. We assume, as before, that the polymer end points are harmonically bound to the center of mass of the polymer.

Initially, we consider an isolated $A-B$ pair. This pair can either (i which case the two centers of mass are linked) or they can be unbound, in which case both the A and and the B and are distributed around their respective center of mass. The distribution is Gaussian. Let us denote the probability that the $A-B$ pair is hybridized by α (the probability that the pair is not hybridized is then $1-\alpha$). The ratio $\alpha/(1-\alpha)$ can be expressed as a ratio of partition functions:

$$\frac{\alpha(R_{AB})}{1-\alpha(R_{AB})} = \frac{Q_{\rm bound}}{Q_{\rm unbound}}, \tag{4.14}$$

where $Q_{\rm bound}$ ($Q_{\rm unbound}$) denotes the intramolecular partition function of the bound (unbound) state. If we use the explicit expressions for these partition functions, we obtain:

$$\begin{aligned}
&\frac{\alpha(R_{AB})}{1-\alpha(R_{AB})} = \\
= \ & \frac{\int dr \exp(-\frac{1}{2}\kappa(r-R_A)^2)\exp(-\frac{1}{2}\kappa(r-R_B)^2) q_{int}(AB)/\Lambda_{AB}^3}{\int dr \exp(-\frac{1}{2}\kappa(r-R_A)^2)\int dr \exp(-\frac{1}{2}\kappa(r-R_B)^2 q_{int}(A)\ q_{int}(B)/(\Lambda_A^3\ \Lambda_B^3)} \\
= \ & \frac{K/\rho_0 \int dr \exp(-\frac{1}{2}\kappa(r-R_A)^2)\exp(-\frac{1}{2}\kappa(r-R_B)^2)}{\int dr \exp(-\frac{1}{2}\kappa(r-R_A)^2)\int dr \exp(-\frac{1}{2}\kappa(r-R_B)^2)} \\
= \ & \frac{K}{\rho_0}\exp\left(-\frac{1}{4}\kappa R_{AB}^2\right)(\pi/\kappa)^{3/2}(\kappa/2\pi)^3 \\
= \ & \frac{K}{\rho_0}\exp(-\frac{1}{4}\kappa R_{AB}^2)(\kappa/4\pi)^{3/2} \equiv f(R_{AB})\ ,
\end{aligned} \tag{4.15}$$

where, in the last two lines, we have used eqn. 4.13 that relates the intramolecular partition functions of the reactants and products to the (known) equilibrium constant K of the dimerization reaction of the free A and B segments. Eqn. 4.15 shows that the dimerization probability depends not only on the equilibrium constant K, but also on distance R_{AB} between the centers of the two polymers chains. defines an effective reaction constant which depends The last line of eqn. 4.15 defines the function $f(R_{AB})$. When $R)_{AB}$, κ and K are given, we can compute $f(R_{AB})$ and thus $\alpha(R_{AB})$:

$$\alpha(R_{AB}) = \frac{f(R_{AB})}{1 + f(R_{AB})}. \tag{4.16}$$

$\alpha(R_{AB})$ allows us to determine the probability of dimerizing two ssDNA of type A and B at a certain distance R_{AB}, assuming that we have only one single dsDNA chain. However, in a solution of coated colloids many DNA strands compete with each other to bind to their complementary strands. In general, there can be many candidates ssDNA of type A that could bind to a given ssDNA of type B. Therefore, we have to compute the partition functions for all possible hybridization states and, from these, determine the relative binding probabilities of all the candidates. The general expression for the probability that end group A will bind to the i-th end group op type B is then of the form:

$$\alpha_i = \frac{K/\rho_0 \exp(-\frac{1}{4}\kappa R^2_{AB_i})(\pi/\kappa)^{3/2}}{\sum_i K/\rho_0 \exp(-\frac{1}{4}\kappa R^2_{AB_i})(\pi/\kappa)^{3/2} + (\kappa/2\pi)^3} \tag{4.17}$$

In the Monte Carlo reaction moves, we first identify all possible hybridization partners of a particular A (or B) segment. The probability that A (resp. B) then binds to the i-th potential partners is then given by α_i. In the next sections we discuss this algorithm in more detail.

4.5.2 Reaction algorithm

Consider a system of chain molecules with "sticky" ends of type A B. A and B can dimerize to form AB. We use the notation:

$$Q_{A_i,B_j} \equiv (\frac{\kappa}{4\pi})^{\frac{3}{2}} \frac{K}{\rho_0} \exp(-\frac{\kappa}{4} R^2_{A_i,B_j}), \tag{4.18}$$

where R_{A_i,B_j} is the distance between particle A_i and particle B_j. We use the following algorithm to allow bonds to from or to break:

1. Choose a reactive end group i at random (bound or unbound).

2. . Calculate Q_{ij} for all possible complementary groups j that can bind to i, and are within a cutoff R_{cut}. If i was initially bound to a specific group j', then j' is included in this list.

3. Calculate S_i as the sum of all the calculated Q_{ij}:

$$S_{ij} = \sum_j Q_{ij}.$$

4. either bind end group i to one of the complementary end groups j with probability $Q_{ij}/(1 + S_i)$ or leave i unbound with a probability $1/(1 + S_i)$.

This algorithm satisfies detailed balance.

For DNA-coated colloid system, once we know the sequence of ssDNA complementary parts we can estimate the hybridization free energy ΔG_{hyb} (see appendix 4.6). K then relates to ΔG_{hyb} through Van't Hoff relation:

$$K = \exp\left(-\frac{\Delta G_{hyb}}{k_{\mathrm{B}}T}\right). \tag{4.19}$$

Eqn. 4.18 implies that two DNA strands have a higher probability to bind when they are close to each other.

4.6 Appendix A : Thermodynamics of DNA base-pairing interactions

Under experimental conditions, the hybridization free energy depends on the properties of the solution, temperature, and most importantly on the sequence and length of the DNA strands. Having the sequence of a double stranded DNA, one can estimate the energetic and entropic cost of melting of the whole strand. Early work of Santalucia [50] provided a unified description of the thermodynamic parameters based on a nearest neighbor (n-n) model. Markham and Zuker [51] developed a freely available package called "Hybrid" based on more up-to-date thermodynamic parameters. The package is available on the DINAMelt web server [52]. The method is commonly used to predict the melting temperature of DNA and RNA strands. Here we briefly discuss the principles of the method:

Consider a hybridized dsDNA, with one strand made of a sequence of the four bases "A", "C", "G", and "T" as monomers. Assume that the other strand has the complementary sequence such that it can bind to the first one without any mismatch

and that a bond has been formed between each pair of complementary bases. Both the energetic cost and the entropic cost (or rather: benefit) of breaking each of these bonds depends on the type of the monomers, their place along the strand, the direction of the strand, and their neighboring monomers. The free energy difference between bound and unbound state of two complementary strands is denoted as:

$$\frac{\Delta G}{RT} = \frac{\Delta H}{RT} - \frac{\Delta S}{R} \tag{4.20}$$

Table 4.1: Unified oligonucleotide ΔH° and ΔS° NN parameters in $1M$ NaCl.

Sequence	ΔH° $\mathrm{k}col/mol$	ΔS° $cal/k^\circ mol$
AA/T	-7.9	-22.2
AT/TA	-7.2	-20.4
TA/AT	-7.2	-21.3
CA/GT	-8.5	-22.7
GT/CA	-8.4	-22.4
CT/GA	-7.8	-21.0
GA/CT	-8.2	-22.2
CG/GC	-10.6	-27.2
GC/CG	-9.8	-24.4
GG/CC	-8.0	-19.9
init. w/term. $G.C$	0.1	-2.8
init. w/term. AT	2.3	4.1
Symmetry correction	0	-1.4

where ΔG is the difference in free energy, ΔH is the difference in enthalpy, and ΔS is the difference in entropy between the hybridized and the random coil state and R is the gas constant. Ref. [50] provides estimates for ΔH° and ΔS°. These values have been determined from melting studies of an appropriately chosen set of dsDNA chains. Table 4.1 shows the value of each parameter at a fixed temperature and in a standard solution. The first column in the table includes all the permutations of two neighboring base-pairs. Note that among all the permutations there are only 10 unique combination of base-pairs next to each other. These are $(5' - 3')^2$ $AA = TT$,

[2] *The polarities of two strands in dsDNA run antiparallel to each other. This polarity is*

$AG = CT$, $AC = GT$, $GA = TC$, $GG = CC$, $TG = CA$, CG, GC, AT, and TA.

For a given sequence of dsDNA one can get the hybridization free energy associated with the complete melting of the strand. Further corrections should be applied regarding changes in temperature and solution. For more information see Ref. [50] and [51]. On DINAMelt web server [52] the melting properties of a given DNA strand can be computed from the sequence information .

The probability of forming a bond between two coated colloids depends not only on ΔG but also on the geometry of the system and on the DNA concentration in the solution. Parameters such as length and stiffness of the DNA strands as well as the shape of the colloidal particles change the effective DNA concentration and the average number of available DNA strands. All these parameters together define an "effective" binding free energy, which might differ largely from the hybridization energy. In simulation studies, this effective binding energy should be determined in order to be able to perform a valid hybridization moves. Different strategies can be taken into account depending on the way we model the coated colloids. For example Licata and Tkachenko [53] presented a statistical- mechanical model for two-particle binding energy for a model of coated-colloids with complementary "rigid" linkers which is then different than our soft picture model for long felxible linkers that has been discussed in section 4.5.1.

indicated by referring to one end as the $3'$ *end and the other as* $5'$ *end [16]*

5 Liquid - vapor transition for DNA coated colloids

5.1 Introduction

In 1957 Alder and Wood demonstrated in numerical simulations that hard spheres could freeze [3, 4]. This work created quite a stir [54] because it showed a crystal phase could have a higher entropy than the liquid phase at the same density. This ran against the conventional wisdom that attraction was needed to cause crystallization of simple liquids. However, Onsager [2] had already shown in the 1940's that a system of slender, hard rods could increase its entropy by undergoing an orientational ordering transition. Since then, many other "entropy-driven" phase transitions have been identified (see, e.g. [55]). For instance, a mixture of hard colloids and non-adsorbing polymers can undergo a liquid-vapor transition because, upon condensation, the loss in entropy of the colloids is compensated by the gain in entropy of the polymers [56, 57]. Yet, this transition is not as surprising as it may seem because, to a first approximation, the polymers induce an attractive pair interaction between the colloids. This makes the mixture of purely repulsive polymers and colloids equivalent to an effective one-component system of colloids with mutual attraction. It is this effective attraction that drives the vapor-liquid condensation. More surprising was a theoretical study by Zilman et al. [58] that showed that a purely entropic condensation transition is possible in a system of inert droplets coated with strongly bound telechelic polymers [59]. In the system studied in Ref. [58], the telechelic polymers could redistribute over the droplets in such a way that they maximized their entropy. In the present chapter, we report numerical simulations that support the mean-field scenario of Ref. [58]. However, there are some differences in the underlying model. Most importantly, as we already discussed in the previous chapter, we consider a

suspension of colloids grafted with long polymers that have "sticky" ends. These polymers can reversibly form bridges between the colloids. In this model, every colloid has a fixed number of linkers. A possible realization of the system that we study is a mixture of colloids, chemically coated with very long strands of double-stranded DNA (dsDNA), functionalized with a single-stranded (ssDNA) linker. These linkers can bind via complementary ssDNA linkers to other DNA-coated colloids [42]. More details on the experiment have been discussed in previous chapter. These systems have the interesting property that individual bonds may be weak but, because of the large number of bonds between two particles, the materials tend to be mechanically strong. Moreover, they should be "self-healing" [60]: if two blobs of a dense network of colloids coated with long DNA are brought into contact, they should be able to reconnect in such a way that the interface disappears. The time scale over which this happens depends on the binding strength between the complementary ssDNA fragments. In this chapter we report our study in two different regime: first in high binding-strength regime where all the DNA sticky ends are hybridized. As we will discuss, a study of this regime allows us to show unambiguously that the phase transition that we observe is entropic in nature. We then extend the model in two ways that make it more realistic. First of all, we consider the situation where the binding free energy for DNA hybridization is finite. This allows us to ascertain to what extent the observed phase behavior is robust. Secondly, we consider the situation where the polymers bind to fixed positions on the surface of the colloids. This is different from the simpler model with infinite binding strengths that we study first, where we have allowed the polymers to be bound to any position on the surface of a colloid. The model with polymer end groups that are mobile on the colloid is easier to study numerically. However, the model with fixed grafting positions is presumably closer to the experimental situation.

5.2 Very high binding energy regime

In what follows, we refer to the particles with complementary functionalization as A and B. Colloids type A and B are identical except that they are coated with complementary strands of ssDNA. We consider the situation where the radius of gyration of the polymer is comparable to the size of the colloids. In Grand Canonical Monte Carlo simulations, the volume, temperature and chemical potential of the system are fixed. As we consider a system where all n polymers per particle are bound to a complementary polymer on another particle, particles are inserted and removed form the system as dimers. The dimers that we attempt to insert are taken from an equilibrated

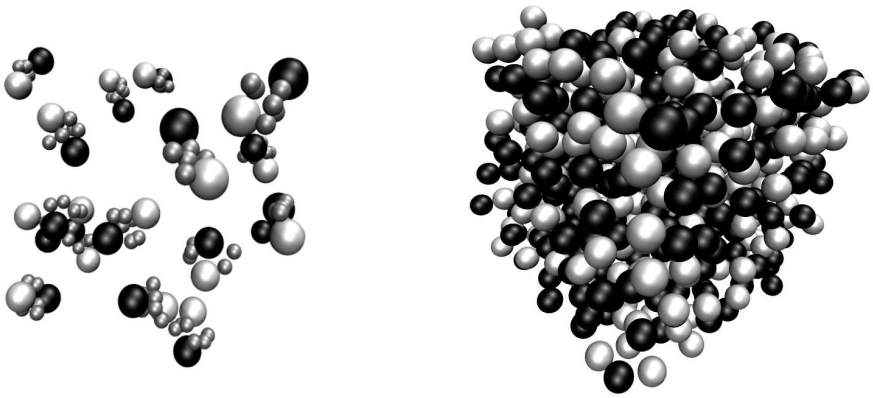


Figure 5.1: Snapshot of a low-density and high density system of a 1:1 mixture of hard colloids (shown as big dark grey and light grey spheres) coated with complementary sticky polymers (shown as small gray spheres in left snapshot)

population of dimers that have intra, but no inter-molecular interactions. When we remove particles, we first check for free dimers (i.e. dimers that are not bound to any other particles). The acceptance rules are constructed such that detailed balance is satisfied. Full details on the method has been discussed in the previous chapter. In our simulations, we considered 1:1 mixtures of hard colloids coated with n sticky polymer arms. We consider the cases $n = 2-10$ — the ratio of the radius of gyration of the polymer to the radius of the colloid was fixed at 0.3, which is a number that can easily be achieved in experiments on DNA-coated colloids [42]. We describe the individual polymer arms as soft colloids, using the approach of Ref. [61, 46]. In particular, the work of Ref. [46] strongly suggests that the soft-colloid picture is to a good approximation transferable to the interaction between polymer moieties that are bound together. In the present work we simply postulate that all polymer-polymer interaction are transferable and are of the form:

$$\frac{V_{p-p}(r)}{k_{\mathrm{B}}T} = 2\exp(-0.7(\frac{r}{R_g})^2) \tag{5.1}$$

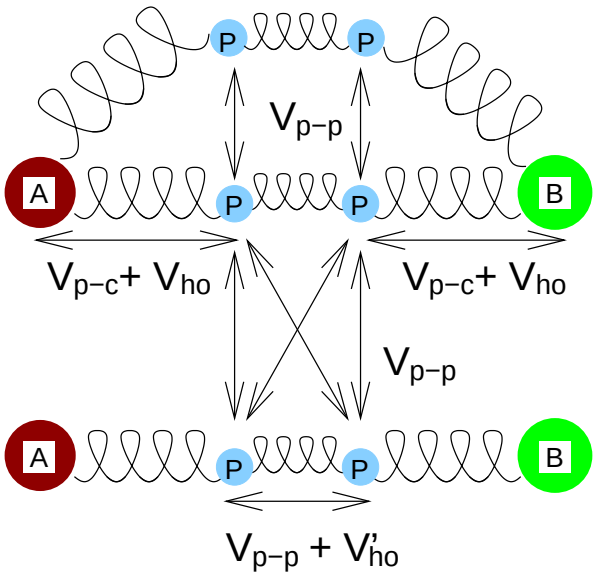


Figure 5.2: Pair-potentials: The transferable repulsive potential between two polymers, V_{p-p}, has the form given by Eqn. 5.1. The transferable potential between polymer and colloid is repulsive and has the form $V_{p-c}(r)$ $580 k_B T \exp(-(r/r_{col})^2/0.23)$. A harmonic spring ($V_{ho}(x) = k_B T x^2/(2R_g^2)$) is assumed to tether the center of mass of each polymer to "its" colloid. When two complementary DNA's are bound, they are assumed to be connected by a harmonic spring (V'_{ho}) that is half as stiff as V_{ho}.

The effective interaction between a free polymer and a hard colloid is not known in closed form [61]. We therefore performed lattice simulation to compute the potential of mean force acting between a self-avoiding polymer and a hard colloid with a size ratio 1:3 (see caption of Fig. 5.2). We found that the repulsive interaction between polymer and colloid is again transferable (i.e. it is the same for bound and unbound polymers). As in Ref. [46], the attractive interaction between a polymer and hard colloid is well described by a simple harmonic spring. The end points of the polymers are considered to be harmonically bound to the center of mass of the soft colloid [46]. One end of the polymer is bound to the surface of a colloid, the other end can bind reversibly to a complementary polymer. The degree of association between the colloids depends on the binding constants between the A and B polymers. To bring out the entropic nature of the phase transitions, we consider the case where the binding constant is very large, implying that there are no free A or B ends. However, we

do assume that bonds can switch, such that the topology of a resulting network can change.

We performed Grand-Canonical Monte Carlo simulations [11] on such a system of functionalized colloids. In our simulations, we chose the thermal energy k_BT as the unit of energy and σ, the diameter of the colloid, as the unit of length. We considered a system with a volume $V = 1000\sigma^3$. In order to speed up the simulation we used cluster moves and cell lists. Cluster moves has been described in details in the appendix 5.5.

Reaction switches The binding-unbinding moves in high hybridization energy regime are slightly different and actually are much easier than the one we described in chapter 4 for a general case. Here in every switch move two pairs of connected polymers, $A_1 - B_1$ and $A_2 - B_2$ will switch partners to lets say $A_1 - B_2$ and $A_2 - B_1$. The value of hybridization energy ΔG_{hyb} (which assumed to be very large) does not appear in the acceptance argument. This is because all the polymers will be always in the form of dimers and therefore through switching move old and new configurations both will have the same number of bonds. The details of the algorithm is presented in appendix 5.4

At very low densities, this system will form a gas of AB dimers. The reason is that, because the polymers are very long, all polymers on an A colloid can bind to a complementary polymer on a B colloid. In contrast to the model considered in Ref. [58], loop formation and, more importantly, polymer fractionation, cannot occur. The resulting dimers are "inert" as they have no unsatisfied bonds. Upon increasing the chemical potential of the system, we find that the colloids undergo a very sudden transition from a low-density "dimer" phase to a uniform, high-density phase where almost all colloids belong to a single, system-spanning cluster (see Fig. 5.1). We stress that, as the degree of association of the polymers is 100%, independent of density, there is no change in the energy of the system upon condensation. The condensation transition is completely driven by the gain in entropy associated with an increase of the number of allowed network topologies. In Fig. 5.3 we plot the density of the system against fugacity of dimers $f_D \equiv \exp(\beta\mu)$. f_D can be interpreted as the density of a gas of non-interacting dimers that is in osmotic equilibrium with the interacting system, and μ is the chemical potential of dimers. Fig. 5.3 shows that at very low densities the dimer concentrations in the system is equal to the fugacity f_D, as it should. The location and magnitude of the density jump depend on the number of polymers per colloid. For the case where we have only two arms per colloids, the magnitude of the density jump is very small (possibly zero): it could not

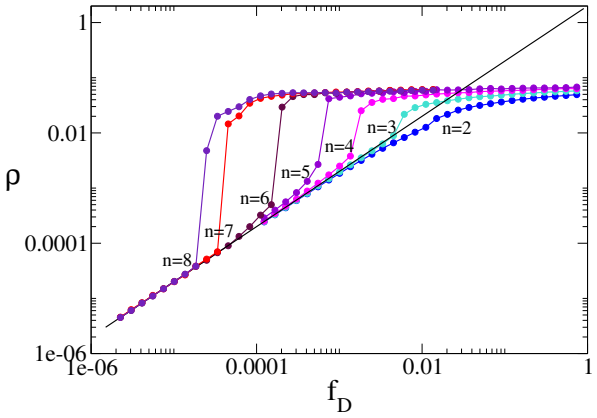


Figure 5.3: Off-lattice grand-canonical simulation; Number density ρ versus the dimer fugacity f_D. The straight line indicates the ideal gas behavior

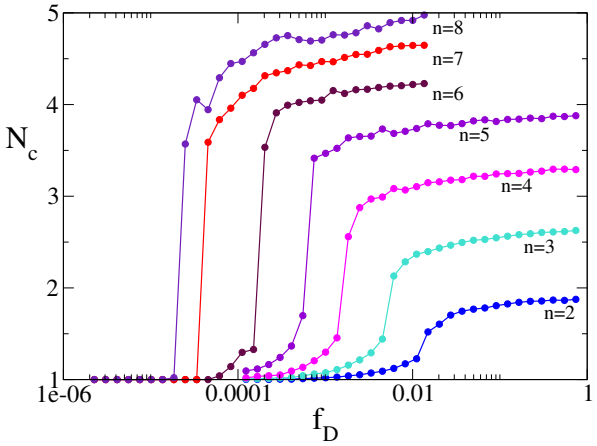


Figure 5.4: Off-lattice grand-canonical simulation; The average number of distinct connected neighbors versus f_D

be resolved in our simulations. It is not surprising that particles with two arms should exhibit qualitatively different behavior because particles with at most two neighbors can only belong to finite linear clusters that are not system-spanning. Nevertheless,

there still is an entropic advantage associated with the formation of a dense phase, as a single configuration can be decorated in many distinct ways by polymer bridges.

The discontinuous transition from vapor to the dense phase is also reflected in the behavior of N_c, the average number of distinct connected neighbors per particle, shown in Fig. 5.4. At low densities, where all colloids form dimers, $N_c = 1$. With increasing chemical potential, N_c undergoes a discontinuous jump, except, possibly, for the case $n = 2$ where there is no system spanning cluster in the dense phase.

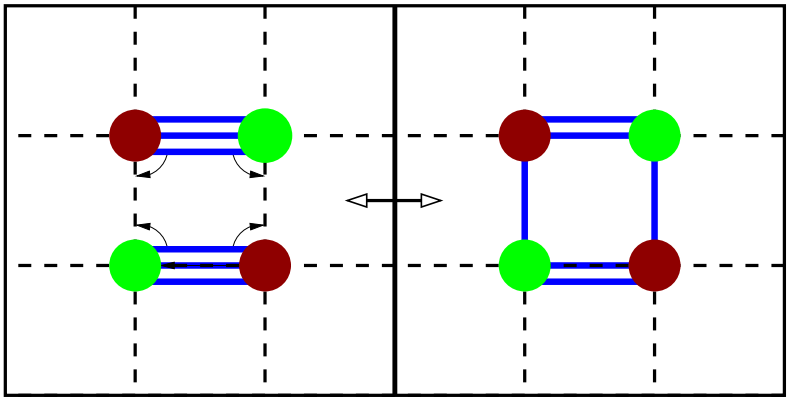


Figure 5.5: Bond switches in lattice simulation.

In order to bring out the essential features of the entropy-driven condensation reaction, we constructed a lattice-gas model of our continuous system. We performed Grand Canonical MC simulations on a simple 3D lattice model that has the following features:

1. A lattice site is either empty, or occupied by an A or a B particle
2. There are equal numbers of A and B particles
3. Every A(B) particle is connected with exactly n bonds to nearest neighbor B(A) particles.

In addition to dimer insertion/removal moves, we perform bond switches that conserve the total number of bonds per particle (see Fig. 5.5). In this way, we ensure that there are no unsatisfied bonds. Fig. 5.6 shows the relation between the occupation of

the lattice and the fugacity of dimers. The behavior shown in Fig. 5.6 is qualitatively very similar to that of the colloid-polymer system. Fig. 5.7 shows N_c as a function of chemical potential for lattice simulations. Similar to the off-lattice simulation, a discontinuous density jump can be seen with increasing chemical potential.

One of the characteristic features of the first order transition is existence of a hysteresis loop. As the lattice simulations are much faster than their off-lattice counterparts, it is easier to probe the hysteretic behavior in lattice simulations. We studied hysteresis by initializing the system in the cluster phase and lowering the chemical potential. We do not observe hysteresis for systems with three bonds per colloid. However, systems with larger number of bonds per colloid, show clear hysteresis loops. This is hardly surprising as networks of multiply-bonded particles are relatively "immune" to the fluctuations that are needed for cluster break-up [62].

We cannot compare our results directly with the mean-field predictions of Ref. [58] because that work considers colloids coated by telechelic polymers where the number of polymers bound to any given colloid is not fixed: due to fractionation, it will be higher in the dense phase than in the dilute phase. In contrast, the linkers in the system studied here cannot fractionate, nor can they loop back onto the same particle. If we consider the limit of the mean-field predictions of Ref. [58] for the case where telechelics can only form bridges and no loops, the phase diagram that is predicted by Ref. [58] shows some similarity to the one that we observe. However, the finite-density dimer phase that we observe before condensation is not found in Ref. [58], presumably because it is preempted by polymer fractionation.

Thus far, our discussion of models for polymer-coated colloid system focused on the case where the polymer endpoint are always bound to complementary polymers. This allowed us to show unambiguously that the condensation transition is of entropic nature. However, in real systems of DNA-coated colloids, the binding strength is finite. Using Grand-Canonical Monte Carlo simulations, we have verified that lowering of the polymer-polymer association constant does not qualitatively change the scenario that we have observed, as long as the binding strength is above a critical value below which the first-order transition disappears. We therefore expect that the entropic condensation should be observable in real DNA-coated colloid systems, provided that the polymers are long enough to allow n of them to bridge the same pair of colloids. Clearly, this condition is not fulfilled for polymers that have a radius of gyration that is much smaller than the radius of the colloids [27]. Coating colloids by long polymers is feasible: Schmatko et al. [42] reported confocal microscopy studies of colloids coated with λ-DNA with a radius of gyration equal to 1.6 times the colloid radius. These experiments observed strong clustering of the colloids, followed by

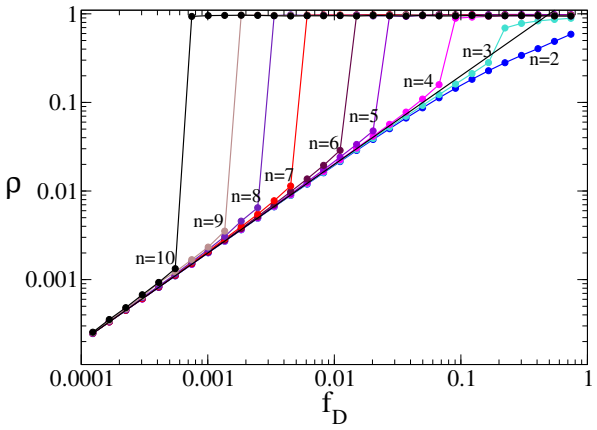


Figure 5.6: Lattice grand-canonical simulation; ρ versus f_D. The straight line indicates the ideal gas behavior.

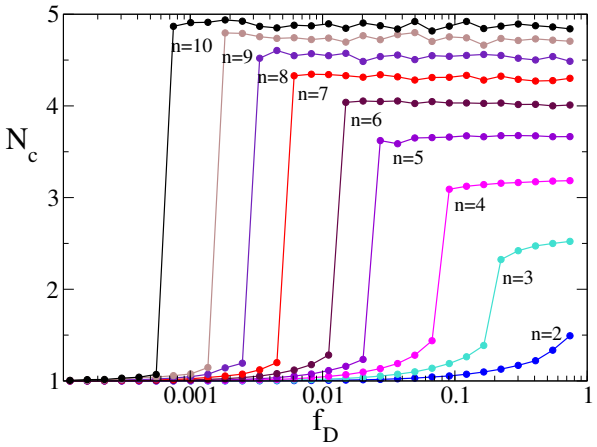


Figure 5.7: Lattice grand-canonical simulation; The average number of distinct connected neighbors versus f_D.

structural arrest. This is due to the strong DNA hybridization under the experimental conditions studied in Ref. [42]. In order to observe the underlying phase transition

predicted here, the experiments would have to be carried out under conditions where the DNA network is able to rearrange on experimental time scales.

5.3 Finite binding strength and fixed grafting points

In the previous section we showed that if the binding strength between complementary ssDNA strands becomes infinite, the system can undergo a first-order gas-liquid transition that is purely entropic in origin. We considered this limit of very high hybridization free energies in order to focus on a situation where the energetic contribution to the binding free energy is the same at all densities. However, in practical situations, the binding free energy for ssDNA hybridization will, of course, be finite and we therefore must determine to what extent the phase behavior that we described in the previous section is changed when we allow bonds to dissociate. In our simulations, the hybridization free energy enters the problem via the expression for the equilibrium constant for the dimerization of two complementary ssDNA strands (eqn. 4.19).

$$K = \exp\left(\frac{\Delta G_{hyb}}{k_{\mathrm{B}}T}\right).$$

Inspection of this equation shows that the quantity that determines the equilibrium constant K is the ratio $\Delta G_{hyb}/(k_{\mathrm{B}}T)$. In other words, a variation of the hybridization free energy may be viewed as an equivalent to a variation of the temperature. It should be stressed that this picture, although convenient, is not quite correct as ΔG_{hyb} contains both an energetic and an entropic part: varying T only changes the ratio $\Delta E_{hyb}/k_{\mathrm{B}}T$. It remains true, however, that variations in T lead to changes in K. Therefore, when we speak about "low" or "high" temperatures, what we mean is large or small values of K. In fact, as mentioned in chapter 4, temperature is one of the important control parameters in experiments on self-assembly of coated particles.

Let us now consider what happens if we start with a situation of infinite binding strength ("$T = 0$") and then lower the equilibrium constant K. Clearly, if K is very small, the DNA-coated colloids will not bind at all and hence the gas-liquid transition will have disappeared. It is therefore logical to assume that, as K decreases, the liquid-vapor transition will shift to higher colloid densities and will eventually move to densities that are unphysical. Alternatively, one might consider the situation where a system with a large equilibrium constant has formed a dense aggregate. Upon heating, this system will reach a point where the dilute phase is more stable and the aggregate will melt spontaneously. How exactly the liquid-vapor transition disappears will depend on the number of DNA chains per colloid. In the present

section we report simulations at finite temperature that allow us to ascertain how sensitive the vapor-liquid phase behavior of DNA-coated colloid is to variations in the binding constant for hybridization. On the basis of these simulations, we can roughly estimate at what value of K the liquid-vapor transition disappears for a given number of DNA chains per colloid. However, as we could not compute the free energy of the dense and dilute phases, we cannot present a real phase diagram.

Our simulation method is similar to the one described in the previous section, except that the way we deal with finite binding constants forces us to use the approach described in chapter 4. As explained above, we consider the case where the DNA chains are grafted at fixed positions on the surface of the colloids.

As before, we model each polymer as a "soft" sphere connected to a hard colloidal sphere, but now the connection point is fixed. This model is expected to be closer to the experimental situation where the biotinilated DNA binds irreversibly to a given streptavidin unit that is grafted to the colloidal surface. In what follows, we refer to the points where DNA chains are attached to the colloidal surface as "patches'. The harmonic free energy of the stretched DNA chain will then be calculated by considering the distance between the center of the soft sphere that represents the polymer to its fixed patch. There are several ways to decorate the colloidal particles with patches: one could either generate a regular patch distribution or one that is random. Here we consider the first case that is a colloid with regular patches on the surface. We chose this situation for the sake of simplicity. However, we stress that it would certainly be interesting to study the (more realistic) case of random patch distributions - in particular when coupled to the case where the number of DNAs per colloid itself is a fluctuating quantity. Such studies are the subject of future work.

As before we refer to the particles with complementary functionalization as A and B. Colloids type A and B are identical except that they are coated with complementary strands of ssDNA. We consider the situation where the radius of gyration of the polymer is comparable to the size of the colloids. In our simulations, we consider 1:1 mixtures of hard colloids coated with n sticky polymer arms. We consider the cases $n = 2, 4, 6,$ and 8. The ratio of the radius of gyration of the polymer to the radius of the colloid was fixed at 0.3, which is a number that can easily be achieved in experiments on DNA-coated colloids [42]. The polymer-polymer interactions were modeled using the soft-sphere pair potentials described in the previous chapter.

The degree of association between the colloids depends on K, the equilibrium constant for hybridization of the stretched complementary ssDNA connected to the A and B colloids. As explained in chapter 4, we can use the available information about binding of isolated ssDNA strands in solution to design a Monte Carlo scheme

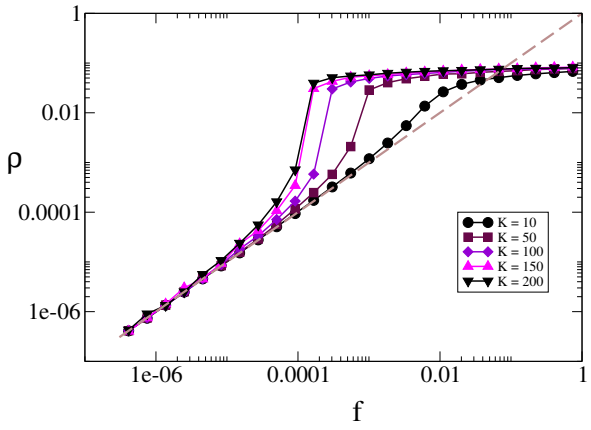


Figure 5.8: Density ρ versus fugacity f for colloids coated with four DNA strands fixed on the surface patches with a tetrahedral arrangement. K denotes the equilibrium constant for hybridization of the ssDNA segments. The transition to the liquid phase gets sharper as K increases.The straight dashed line indicates the ideal gas behavior.

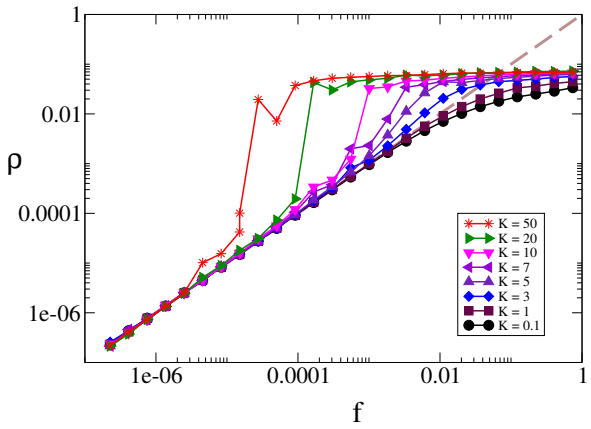


Figure 5.9: Density ρ versus fugacity f for colloids coated with six DNA strands fixed on the surface patches arranged on the vertices of an octahedron. The straight dashed line indicates the ideal gas behavior. K denotes the equilibrium constant for hybridization of the ssDNA segments. The transition to the liquid phase gets sharper as K increases.

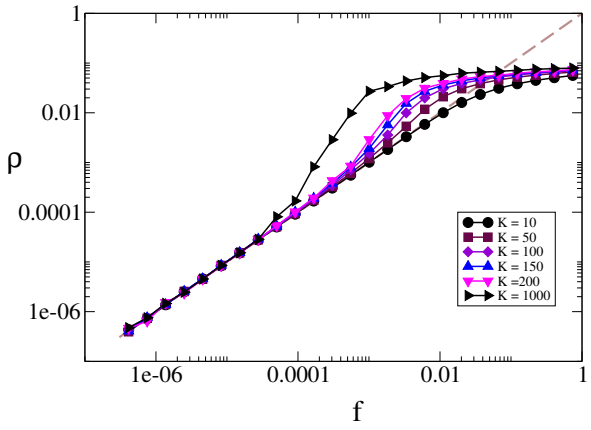


Figure 5.10: Density ρ versus fugacity f for colloids coated with two DNA strands fixed on the surface patches opposite to each other. The straight dashed line indicates the ideal gas behavior.K denotes the equilibrium constant for hybridization of the ssDNA segments. The transition to the liquid state remains smooth even for the highest values of K that we consider. The changeover in slope that we observe is mainly caused by increased dimerization as the density of the gas is increased.

that will allow long, ssDNA-terminated dsDNA chains to bind to, or unbind from, a chain with complementary "sticky" end. A large value of K corresponds to a large hybridization free energy and thus approaches the conditions of 100% hybridization that we studied in the previous section.

As the DNAs are grafted to fixed patches on the colloidal surface, our MC algorithm should not only handle the translational motion of the colloids, but also their rotation. A well-known and convenient way to implement arbitrary rotations of a 3D object is to use a quaternion-based description (see appendix 5.6). Rotational moves are essential as the distance between two patches on neighboring colloids depends both on the center-of-mass separation and the orientation of the colloids. Clearly, two polymers that are grafted to facing sides of the colloids will bind more easily (i.e. will have to stretch less in order to bind) than polymers that are on the far side. Apart from the rotational moves, we use the same methods as before: cell lists (see, e.g. Ref. [11]) and cluster moves (see appendix 5.5).

The aim of the present section to show qualitatively how a finite binding constant affects the phase behavior reported in the previous section. To this end we performed

grand-canonical simulations where we determined the density of the system as a function of the fugacity of the colloids, for a range of values of the equilibrium constant K. Fig. 5.8 shows a plot of density versus fugacity for colloids with four polymeric arms. Note that the fugacity f that we use here is the fugacity of colloidal particles. In the previous section, where we considered the limit $K \to \infty$, the low-density limit of the system was an ideal gas of dimers. However, for finite K, dimers will dissociate at low densities and the limiting state is an ideal gas of isolated A and B colloids. For the case $n = 4$ the DNA polymers were attached to patches that were arranged on the corners of a tetrahedron. Note that for low values of K (e.g. K=10), the density in the system varies continuously with fugacity. As K increases, the variation in density becomes sharper and for $K \geq 100$ the behavior is very similar to that observed in the limit $K \to \infty$.

Fig. 5.9 shows a similar plot as shown in Fig. 5.8, but now for six polymers per colloid. Compared to the case $n = 4$, the transition already becomes sharp at lower hybridization free energies. Fig. 5.10 shows the same behavior for the case where we have two DNA strands per colloid. Here we can see that the density profile develops a shoulder and the slope of the profile increases as we increase K . However even for the highest K values the profile remains smooth and shows no indication of a first-order phase transition - again in agreement with the behavior of the $K \to \infty$ limit discussed in the previous section.

As we cannot directly determine the point where the phase transition becomes first order, we use an indirect method to estimate the values of n and K where this happens. To this end, we compute for various values of n the inverse of the maximum slope of $(d \ln \rho / d \ln f)$. In other words, we compute the minimum of

$$D \equiv \frac{d \ln f}{d \ln \rho} = \frac{dP}{d\rho}.$$

For an ideal gas, $D = 1$ (we use units where $k_{\mathrm{B}}T = 1$). At a critical point, $(dP/d\rho)$ vanishes, and hence we can estimate the location of the critical point by plotting $(dP/d\rho)$ vs $\ln K$. The reason for plotting D versus $\ln K$ is that the latter quantity is proportional to $1/T$. As most of our points are not very close to the critical point, we can safely ignore possible non-classical behavior in the vicinity of the critical point and simply assume that D vanishes linearly as we approach $\ln K_c$. Unfortunately, our present data only allow a very rough estimate of the location of the critical point. This is because our discrete data and rather large fugacity spacing lead to large error bars in the slopes.

At high K, the slopes of the curves for large n converge to a value that is slightly larger than zero. In contrast, in the limit of small K the slopes converge to one.

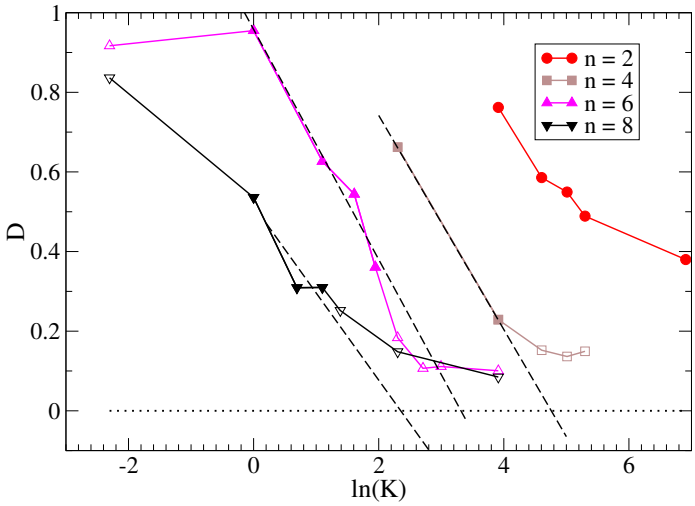


Figure 5.11: D, defined as the inverse of maximum slope of $\ln \rho$ versus $\ln f$, as a function of $\ln K$. High K corresponds to a large hybridization free energy. As explained in the text, the values corresponding to the open symbols are not reliable. These points have therefore been ignored in the rough fit. The dashed lines show the linear extrapolation to the part of the curves shown in closed symbols. K_c is the value of K when the dashed lines meet line $D = 0$. The curves converge to a value which is slightly larger than zero. This is because of our discrete data and our rather large value for fugacity intervals.

Ideally, a linear extrapolation of D vs $\ln K$ would reach a value of zero at a finite value of K. However, it is hard to extract the linear regions in Fig. 5.11, since we have only few points between the "low" and the "high" K regions. To obtain a rough estimate of the critical value of K_c, we fitted a line to the D points in Fig. 5.11, in the intermediate regime. As n increases, the estimate of K_c moves to lower values of K_c (see Fig. 5.12), or, using our loose analogy between $\ln K_c$ and $1/T$, to higher temperatures. This observation indicates that the phase transition that we observe in the limit $K \to \infty$ is robust, in the sense that it survives also for finite K. However, below a certain (n-dependent) value of K, the gas-liquid transition disappears. In particular, as K_c must be positive and finite, $1/K_c \geq 0$. Fig. 5.12 suggests that for $n \leq 2$, the estimated value of $1/K_c$ becomes negative. This suggests that no first-order phase transitions are possible for $n \leq 2$.

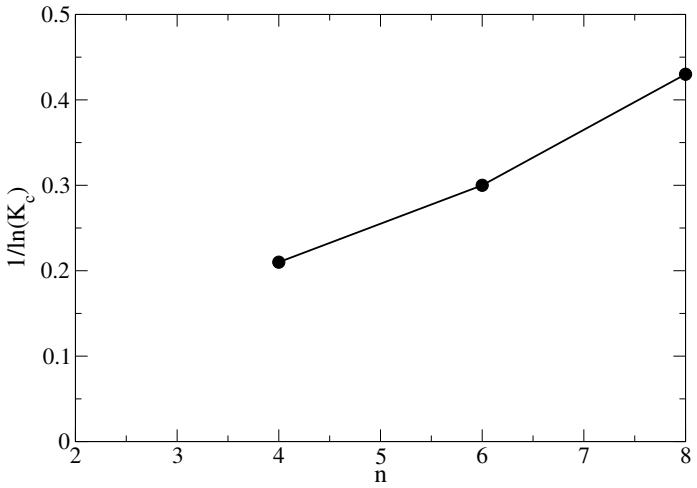


Figure 5.12: The estimate of $1/\ln K_c$ versus number of arms n. As n increses $1/\ln K_c$ goes to higher values.

5.4 Appendix A: Reaction switches

For large values of the hybridization equilibrium constant and, a fortiori, in the limit $K \to \infty$, special Monte Carlo moves are required to change the topology of the network of DNA "bonds" linking different colloids. To achieve this, we employ the following algorithm.

1. Randomly choose a position r in the simulation box.
2. Make a list of all the pairs of connected polymers with centers-of-mass within a range r_{cut} from r. If not at least two A-B pairs are located within the cutoff volume, the trial move is rejected.
3. Randomly choose two of the pairs from the list of pairs within the cutoff volume. Let us denote these pairs as $A_1 - B_1$ and $A_2 - B_2$
4. Calculate the sum of $U_{\rm old}$s for these two pairs, where U is given by Eqn. 5.2 below:

$$U_{A_i,B_j} = \frac{\kappa}{4} R^2_{A_i,B_j}. \tag{5.2}$$

5. Attempt to switch the bonds between these particles: the trial configuration in our example will be $A_1 - B_2$ and $A_2 - B_1$; calculate the sum of $U_{\rm new}$.

6. Accept the move with probability $\exp(-\beta\Delta U)$ where $\Delta U = U_{\rm new} - U_{\rm old}$.

note: The cutoff distance r_{cut} is fixed at the beginning of the simulation and it does not change. Introducing r_{cut} is not necessary however it speeds up the simulation. Ideally, one cutoff volume contains on average two bound pairs.

note: During a bond-switch move, the hybridization free energy does not play a role because the number of bonds before and after the move is the same. Therefore, all the constant coefficients $K/[\rho_0(\kappa/(4\pi))^{3/2}]$ drop out of the acceptance probability.

5.5 Appendix B: Cluster moves

If one performs a simulation of a system consisting of composite objects, it may be quite inefficient to use only conventional MC moves that attempt to move the entire object by displacing one of the constituent units at a time. For such systems, it is often advisable to combine individual particle moves with cluster moves that attempt to displace the composite object as a whole. If the composite object is a cluster of particles, some care should be taken to define the cluster in a way that will not violate detailed balance (see e.g. [11]). However, for the case discussed in the present chapter, the definition of clustering is straightforward as we have DNA bonds between the particles that constitute a cluster. Normally a cluster of bonded colloids is a fairly compact object. In a cluster move, we first label all particles according to the cluster that they belong to. We then choose one of the colloids with equal probability and attempt to displace the cluster to which the selected particle belongs in an otherwise normal MC move. To compute the acceptance probability of such a cluster move we must calculate the sum of all the pair-interactions between particles inside the cluster and all other particles except the particles that belong to the same cluster as, during a cluster move, all relative distances within a cluster are conserved.

5.6 Appendix C: Quaternions

Here we briefly explain what are quaternions and how they can be used to rotate a vector in space. In general a quaternion q can be represented with a complex number

with three different imaginary parts:

$$q = q_0 + iq_1 + jq_2 + kq_3, \tag{5.3}$$

and the only main rules between imaginary units i, j, and k are:.

$$i^2 = j^2 = k^2 = ijk = -1. \tag{5.4}$$

A quaternion can also be represented as vector in imaginary space plus a scalar. In other words a vector is a quaternion with zero "real" part. Now consider a vector $\vec{v}$ which is represented by $\vec{v} = (0, \mathbf{v})$ and consider the transformation

$$\vec{v}' = q\vec{v}q^* \tag{5.5}$$

where q is a unit quaternion and therefore

$$q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1. \tag{5.6}$$

and $q^* = (q_0, -\mathbf{q})$ is the conjugate of $q = (q_0, \mathbf{q})$. It is straightforward to prove that such a transformation preserves lengths and angles and the triple product, and therefore represents a rotation. The rotation will be around the unit axis of vector ω that is defined by the vector part of the quaternion

$$\omega = (q_1, q_2, q_3) \tag{5.7}$$

and the angle of rotation will be specified by the following equation

$$\cos(\theta) = 2q_0^2 - 1 \tag{5.8}$$

Therefore any quaternion $q = (q_0, \mathbf{q})$ will uniquely define a rotation through an angle θ around axis ω which is given in the above equations. Refs. [63, 64] provide more information on proof of the above arguments, quaternions in general, and their applications.

To rotate a colloid, we employ the following algorithm:

1. Randomly choose a colloid and generate a random quaternion q that satisfies Eqn. 5.6.

2. Calculate the sum of U_{old} where U includes the spring interactions between the chosen colloid and its bond polymers.

3. Rotate all the patches of the chosen colloid around colloid's CM, using Eqn. 5.5.

4. Calculate the sum of U_{new}.

5. Accept the move with probability $\exp(-\beta\Delta U)$ where $\Delta U = U_{\mathrm{new}} - U_{\mathrm{old}}$.

Bibliography

[1] de Gennes PG (1988) *Soft Matter (Nobel Lecture).* Angewandte Chemie International Edition in English **31**:842 – 845.

[2] Onsager L (1949) *The effects of shape on the interaction of colloid particles.* Ann NY Acad Sci **51**:627.

[3] Alder BJ, Wainwright TE (1957) *Studies in molecular dynamics. J Chem Phys* **27**:1208.

[4] Wood WW, Jacobson JD (1957) *Preliminary results from a recalculation of the monte carlo equation of state of hard spheres.* J Chem Phys **27**:1207.

[5] Nykypanchuk D, Maye M, van der Lelie D, Gang O (2008) *DNA-guided crystallization of colloidal nanoparticles.* Nature **451**:549–552.

[6] Park SY, Lytton-Jean AKR, Lee B, Weigand S, Schatz GC, et al. (2008) *DNA-programmable nanoparticle crystallization .* Nature **451**:553–556.

[7] Jin R, Wu G, Li Z, Mirkin CA, Schatz GC (2003) *What controls the melting properties of DNA-linked gold nanoparticle assemblies? . J Am Chem Soc* **125**:1643.

[8] Rogers PH, Michel E, Bauer CA, Vanderet S, Hansen D, et al. (2005) *Selective, controllable, and reversible aggregation of polystyrene latex microspheres via DNA hybridization .* Langmuir **21**:5562.

[9] Valignat MP, Theodoly O, Crocker JC, Russel WB, Chaikin PM (2005) *Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids.* Proc Natl Acad Sci USA **102**:4225.

[10] Kim AJ, Biancaniello PL, Crocker JC (2006) *Engineering DNA-mediated colloidal crystallization.* Langmuir **22**:1991–2001.

[11] Frenkel D, Smit B (2002) *Understanding molecular simulations: from algorithms to applications.* San Diego: Academic Press.

[12] Ladd AJC (1997) *Sedimentation of homogeneous suspensions of non-Brownian spheres.* *Physics of Fluids* **9**:491–499.

[13] Malevanets A, Kapral R (1999) *Mesoscopic model for solvent dynamics.* *Journal of Chemical Physics* **110**:8605–8613.

[14] Brady JF, Bossis G (1988) *Stokesian dynamics.* *Annual Review of Fluid Mechanics* **20**:111–157.

[15] de Gennes PG (1988) *Scaling concepts in polymer physic.* Ithaca: Cornell University Press.

[16] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, et al. (2002) *Molecular Biology of The Cell 4th Ed.* 29 West 35th Street. New York, NY 10001-2299: Garland Science Taylor & Francis Group.

[17] Kremer K, Binder K (1988) *Monte Carlo Simulations of Lattice Models for Macromolecules.* *Comp Phys Rep* **7**:259.

[18] Binder K (1995) *Monte Carlo and Molecular Dynamic Simulations in Polymer Science.* New York: Oxford University Press.

[19] Sokal AD (1995) *Monte Carlo Methods for the Self-avoiding Walk* In: *Monte Carlo and Molecular Dynamics Simulations in Polymer Science.* New York: Oxford University Press.

[20] Karayiannis NC, Mavrantzas VG, Theodorou DN (2002) *A novel Monte Carlo scheme for the rapid equilibration of atomistic model polymer systems of precisely defined molecular architecture.* *Phys Rev Lett* **88**:105503.

[21] Yan QL, De Pablo JJ (1999) *Hyper-parallel tempering Monte Carlo: Application to the Lennard-Jones fluid and the restricted primitive model.* *J Chem Phys* **111**:9509.

[22] Houdayer J (2002) *The wormhole move: A new algorithm for polymer simulations* . *J Chem Phys* **116**:1783.

[23] Mooij GCAM, Frenkel D (1991) *Novel scheme to compute chemical-potentials of chain molecules on a lattice.* *Mol Phys* **74**:41.

[24] Bolhuis PG, Louis AA, Hansen JP (2001) *Accurate effective pair potentials for polymer solutions.* *J Chem Phys* **114**:4296.

[25] Tuinier R, Aarts D, Wensink HH, Lekkerkerker HNW (2003) *Pair interaction and phase separation in mixtures of colloids and excluded volume polymers.* *Phys Chem Chem Phys* **5**:3707.

[26] Nickel BG (1991) *One-parameter recursion model for flexible-chain-polymers.* *Macromolecules* **24**:1358.

[27] Biancaniello PL, Kim AJ, Crocker JC (2005) *Colloidal interactions and self-assembly using DNA hybridization.* *Phys Rev Lett* **94**:058302.

[28] Lawrence CJ, Dawe RK, Christie KR, Cleveland DW, Dawson SC, et al. (2004) *A standardized kinesin nomenclature.* *J Cell Biol* **167**:19.

[29] Carter BC, Gross SP (2005) *Mechanics of the kinesin step* . *Nature* **435**:308–312.

[30] Kerssemakers JW, Muntenu EL, Laan L, Noetzel TL, Janson ME, et al. (2006) *Assembly dynamics of microtubules at molecular resolution.* *Nature* **442**:709.

[31] Hua W, Young EC, Fleming ML, Gelles J (1997) *Coupling of kinesin steps to ATP hydrolysis.* *Nature* **388**:390.

[32] Milescu LS, Yildiz A, Selvin PR, Sachs F (2006) *Maximum likelihood estimation of molecular motor kinetics from staircase dwell-time sequences.* *Biophys J* **91**:1156.

[33] Frenkel D, Ernst MH (1989) *Simulation of diffusion in a two-dimentional lattice-gas cellular automaton: A test of mode-coupling theory* . *Phys Rev Lett* **63**:2165–2168.

[34] Meijer EJ, Frenkel D (1991) *Computer simulation of polymer-induced clustering of colloids* . *Phys Rev Lett* **67**:1110–1113.

[35] Bozorgui B, Frenkel D (2007) *Lattice-Based Monte Carlo method for thelechelic chain molecules* . *Phys Rev E* **75**:036708.

[36] Sadler BM, Swami A (1998) *Analysis of wavelet transform multiscale products for step detection and estimation.* *Army Research Laboratory Tech Rep ARL-TR- 1664* .

[37] Sadler BM, Swami A (1999) *Analysis of multiscale products for step detection and estimation.* **IEEE Trans Inform Theory** **45**:1043–1051.

[38] Kalafut B, Visscher K (2008) *An objective, model-independent method for detection of non-uniform steps in noisy signals* . **Computer Physics Communications Article in Press, doi:101016/jcpc200806008** .

[39] Carter BC, Vershinin M, Gross SP (2008) *A Comparison of Step-Detection Methods: How Well Can You Do?* **Biophys J** **94**:306–319.

[40] Watson JD, Crick FHC (1953) *Molecular structure of nucleic acids - A structure for deoxyribose nucleic acid.* **Nature** **171**:737.

[41] Geerts N, Schmatko T, Eiser E (2008) *Clustering versus percolation in the assembly of colloids coated with long DNA* . **Langmuir** **24**:5118–5123.

[42] Schmatko T, Bozorgui B, Geerts N, Frenkel D, Eiser E, et al. (2007) *A finite-cluster phase in lambda-DNA-coated colloids.* **Soft Matter** **3**:703.

[43] Lukatsky D, Frenkel D (2005) *Surface and bulk dissolution properties, and selectivity of DNA-linked nanoparticle assemblies* . **J Chen Phys** **122**:214904.

[44] Tkachenko AV (2002) *Morphological diversity of DNA-colloidal self-assembly.* **Phys Rev Lett** **89**:148303.

[45] Bhatia SR, Russel WB (2000) *End-capped associative polymer chains between nanospheres: attraction in ideal solutions.* **Macromolecules** **33**:5713.

[46] Pierleoni C, Addison A, Hansen JP, Krakoviack V (2006) *multiscale coarse graining of diblock copolymer self-assembly: From monomers to ordered micelles.* **Phys Rev Lett** **96**:128302.

[47] Safran SA (1994) *Statistical thermodynamics of surfaces, interfaces, and membranes.* Reading: Addison Wesley.

[48] Pelissetto A, Hansen JP (2006) *An effective two-component description of colloid-polymer phase separation.* **Macromolecules** **39**:9571–9580.

[49] Rubenstein M, Colby RH (2003) *Polymer Physics.* New York: Oxford University Press.

[50] SantaLucia JR (1998) *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics* . *PNAS* **95**:1460–1465.

[51] Markham NR, Zuker M (2005) *DINAMelt web server for nucleic acid melting prediction.* *Nucleic Acid Research* **33**:W577–W581.

[52] See http://dinamelt.bioinfo.rpi.edu/twostate-fold.php.

[53] Licata NA, Tkachenko AV (2006) *Statistical mechanics of DNA-mediated colloidal aggregation* . *Phys Rev E* **74**:041408.

[54] Jerome K, Percus ED (1963) *The Many-Body Problem.* New York: Interscience.

[55] Frenkel D (1999) *Entropy-driven phase transitions.* *Physica A* **263**:26.

[56] Asakura S, Oosawa F (1958) *Interaction between particles suspended in solutions of macromolecules.* *J Pol Sci* **33**:183.

[57] Vrij A (1976) *Polymers at interfaces and the interactions in colloidal dispersions.* *Pure Appl Chem* **48**:471.

[58] Zilman A, Kieffer J, Molino F, Porte G, Safran SA (2003) *Entropic phase separation in polymer-microemulsion networks.* *Phys Rev Lett* **91**:015901.

[59] Michel E, Appell J, Molino F, Kieffer J, Porte G (2001) *Unstable flow and nonmonotonic flow curves of transient networks.* *J Rheol (NY)* **45**:1465.

[60] Cordier P, Tournilhac F, Soulié-Ziakovic C, Leibler L (2008) *Self-healing and thermoreversible rubber from supramolecular assembly.* *Nature* **451**:977.

[61] Bolhuis PG, Louis AA (2002) *How to derive and parameterize effective potentials in colloid-polymer mixtures.* *Macromolecules* **35**:1860.

[62] Hed G, Safran SA (2006) *The immunity of polymer-microemulsion networks* . *Eur Phys J* **19**:69.

[63] See `http://www.cis.udel.edu/ chandra/849/Fall04/Talks/absolute-orientation-handout.pdf`.

[64] See `http://ai.stanford.edu/diebel/attitude.html`.

Summary

This thesis focuses on the use of simulations to study the behavior of colloidal particles that form multiple "bonds" via chain molecules with "sticky" ends. These "telechelic" chains can either bind reversibly to the surface of the colloids or to the complementary grafted polymers (in particular DNA) bound to other colloids. We provide a general introduction on the topic in Chapter 1, and we introduce some of the terms that we frequently use in the thesis.

In chapter 2, we describe a Grand-Canonical Monte-Carlo technique for lattice polymers that is particularly efficient for simulations of "telechelic" polymers that interact strongly with surfaces. The combination of high binding strength and excluded volume interactions between the polymers makes "brute force" calculations difficult. For this situation, we propose an algorithm that combines the so-called "moment-propagation (MP)" with the configuration bias method. The basic idea of our method is to use the statistics of non-self-avoiding walks to grow self-avoiding walks (SAW) on a lattice. We show that it is feasible to implement the MP scheme in such a way that detailed balance is satisfied. In order to test the efficiency of the MP method, we specifically apply it to two sets of problems. In the first part, we compute the depletion interaction between two plates and compare results with existing data [25]. The results show that the MP scheme quantitatively reproduces the known behavior of this model system. However, the MP approach is less efficient than configurational bias Monte Carlo (CBMC) in this case. The situation is reversed in the case of telechelic polymers between two plates. In the second part, we study the telechelic chains and we show that, in particular in the case of highly heterogeneous surfaces, the MP method is more efficient than CBMC. Moreover, the relative advantage of MP is more pronounced at high volume fractions.

The "moment-propagation (MP)" methodology developed in chapter 2 can be applied in a completely different part of biophysics: we shall argue that the moment-propagation method offers a novel, and uniquely sensitive, method to identify underlying steps in noisy experimental data on the motion of molecular motors along linear "tracks" (e.g. kinesin on microtubules). This "detour" is the subject of Chapter

3. The key observation that underlies the approach that we propose here is the following: what we aim to extract is not the precise trajectory of the molecular motor, but the parameters (step-size(s) and step frequency) that are intrinsic properties of the molecular motor under the experimental conditions. There may be many step histories that fit the experimental data to within the experimental error and we argue that it is often better to extract the desired information from the set of possible stepping-motor trajectories. We will exploit the analogy between stepping-motor trajectories and realizations of a directed random walk. The algorithm that we propose in chapter 3 is based on the numerical estimate of the "free energy" of an ensemble of directed random walks with different step sizes. We argue that this method allows us to detect the underlying step size of a molecular motor under "noisy" conditions where existing methods fail. We show that the method is capable of detecting a single step size at several times lower signal-to-noise ratio than the existing methods.

Chapters 4 and 5 contain a description of the coarse-grained model that we use to simulate systems of colloidal particles coated with long DNA strands with "sticky" ends. The sticky ends are "complementary" ssDNA strands that can selectively bind to each other. In particular, chapter 4 describes a Grand Canonical MC simulation technique to study the self-assembly of mixtures of colloids that are coated with "sticky" DNA strands. Using a coarse-grained model to represent DNA strand as "soft" spheres, we study the effective interactions between the colloids. We use lattice Monte Carlo simulations as well as existing studies on colloid-polymer pair-potentials to address the problem of effective interactions between a telechelic polymer and a colloid at zero density. Moreover, We develop a method that allows us to relate the binding properties of DNA-coated colloids to the (known) equilibrium constants for dimerization in solutions of the isolated "sticky" ssDNA strands. Based on this method, we build our binding - unbinding algorithm for the sticky strands.

In chapter 5 we use the GC method mentioned above to simulate an equimolar mixture of hard colloids coated with "long" polymers with complementary ssDNA strands ends. In the first part, we study the regime where the ends are strongly associated. In this regime there are no "free" sticky ends in the system and we use Monte Carlo "switch" moves that allow for rearrangement of bonds in pairs. Under these conditions we observe a first order vapor-liquid transition from a dilute gas of colloidal dimers to a dense, liquid-like phase. This transition is driven exclusively by the increase in entropy associated with bond disorder. In the second part of the chapter we show qualitatively how a finite binding constant affects the phase behavior. Here we use the GC method together with the binding - unbinding moves that has been introduced in chapter 4. We show that the transition persists as we switch to low but

finite temperatures. We study how the critical point at which the first-order transition disappears depends on the number ratio between polymers and colloids.

Samenvatting[1]

De focus van dit proefschrift ligt op het gebruik van computersimulaties van colloïdale deeltjes. Deze deeltjes kunnen onderling meerdere bindingen aangaan door middel van hun ketenmoleculen met "plakkerige" uiteinden. Deze "telechelische" ketens kunnen zowel aan het oppervlak van de deeltjes als aan de com plementaire ketenuiteinden van de overige deeltjes binden (voornamelijk in het geval van DNA). In hoofdstuk 1 geven we een algemene introduct ie over dit onderwerp, ook worden enkele veelgebruikte termen binnen dit proefschrift geïntroduceerd.

In hoofdtuk 2, beschrijven we een Groot-Canonische Monte-Carlo techniek voor polymeren op een rooster. Deze techniek is bij uitstek geschikt voor de simulatie van telechelische polymeren die een sterke interactie met oppervlaktes hebben. De combinatie van de hoge bindingssterkte en uitgesloten volume interacties tussen de polymeren maakt "brute kracht" berekingen moeilijk. Voor deze situatie, stellen wij een algoritme voor, die de zogenoemde "momental e propagatie" (MP) met de configuratie-bias methode combineerd. De basis gedachte achter onze methode is om de statistieken van niet-zelfmijdende wandeling en te gebruiken om zelfmijdende wandelingen op het raster werk te laten ontstaan. We laten zien dat het mogelijk is het MP schema zo te implementeren dat aa n de gedetailleerde balans wordt voldaan. Om de efficientie van de MP methode te testen, passen wij deze toe op twee vraagstukken. In het eerste vraagstuk, berekenen we de depletie-interactie tussen twee platen en vergelijken deze met resultaten uit al bestaande data [25]. De resultaten laten zien dat het MP schema quantitatief het gedrag van dit model systeem reproduceerd. Hoewel, in dit geval deze methode minder presteert dan de configuratie-bias Monte -Carlo (CBMC). Het omgekeerde is het geval als telechelische polymeren aanwezig zijn tussen de platen. In het tweede vraagstuk, bestuderen we telechelische ketens en laten we zien dat, vooral in het geval van zeer hetrogene oppervlaktes, de MP methode efficienter is dan de CBMC. Bovendien is het relatieve gewin van MP meer in het oog springend bij hoge volume fracties.

De door ons ontwikkelde "momentale propagatie" (MP) methodologie uit hoofd-

[1] Translation by Nienke Geerts

stuk 2 kan ook toegepast worden in compleet andere delen van de biofy sica. We hebben sterke argumenten om aan te tonen dat de MP methode een nieuwe, unieke, zeer gevoelige mogelijkheid biedt om ondeliggende stappen te identificeren in met ruis omgeven experimentele data van de voortbeweging van moleculaire motors over lineaire "wegen" (bv kinesine op microtubuli). Dit zijproject is het onderwerp van hoofdstuk 3. Het sterkste punt van onze aanpak is dat we niet de precieze weg die de motor heeft afgelegd uit de data proberen te halen, maar de parameters (stapgrootte en stapfrequentie) die de intrinsieke eigenschappen van de motor onder bepaalde experimentele condities zijn. Aan gezien er binnen de experimentele fout vele stap geschiedenissen kunnen zijn die op de experimentele data passen, stellen wij dat het vaak beter is om de gew ilde informatie uit alle mogelijke afgelegde wegen van de stappende-motor te halen. Hierbij zullen wij de gelijksoortigheid tussen de afgelegde weg van de stappende motor en de realisatie van een gerichte dronkemanswandeling uit buiten. Het algoritme wat we in hoofdstuk 3 voorstellen is gebaseerd o p een nummerieke schatting van de "vrije energie" van een ensemble gegenereerd door gerichte dronkemanswandelingen met verschillende stapgrootte. We bearg umenteren dat deze methode ons instaat stelt om de onderliggende stapgrootte van een moleculaire motor te achterhalen uit data met ruis, waar andere methode s falen. Ook laten we zien dat de methode in staat is de grootte van een stap te bepalen bij een vele malen kleinere signaal-tot-ruis verhouding dan de bes taande methodes.

Hoofdstuk 4 en 5 bevatten een beschrijving van een "grofkorrelig" model, dat gebruikt wordt om colloïdale deeltjes me t lange DNA ketens met "plakkerige" uiteinden te simuleren. De uiteinden zijn complementaire ssDNA stukjes, die selectief aan elkaar kunnen binden. Hoofds tuk 4, in het bijzonder, beschrijft een Groot-Canonische Monte-Carlo simulatie techniek om de zelfbindendheid van dit soort deeltjes te bestu deren. Door gebruik te maken van een grofkorrelig model dat DNA ketens vertaald naar "zachte" bollen, bestuderen we de effectieve interactie tussen de dee ltjes. We gebruiken zowel Monte-Carlo rasterwerksimulaties als bestaande studies over colloïdaal deeltje-polymeer paarpotentialen om het probleem van effect ieve interacties tussen een telechelisch polymeer en een colloïdaal deeltje bij oneindig lage druk aan te geven. Ook hebben we een methode ontwikkeld om de bindingseigenschappen van DNA gedeklaagde deeltjes te relateren aan de (bekende) evenwichtsconstante van het vormen van dimeren onder geïsoleerde "plakkeri ge" ssDNA ketens in oplossing. Op deze methode is ons binding-ontbindingsalgoritme voor de plakkerige uiteindes gebaseerd.

In hoofdstuk 5 gebruiken we de bovengenoemde GC-methode om een mix van gelijke hoeveelheid harde deeltjes met "lange polymere" arm en met complemtaire

ssDNA uiteinden te simuleren. Eerst wordt het regime onderzocht waarbij de uiteinden sterk met elkaar associëren. In dit regime zijn erg geen ongebonden uiteinden binnen het systeem.

Acknowledgments

This book is the result of the four years PhD program that I did in AMOLF, during which I learned extensively from daily discussions and collaborations. I gathered many unforgettable memories of friendship and fun. First I would like to thank my supervisor, Daan Frenkel, for many helpful discussions and for the time he spent even when he was extremely busy. Daan, thank you very much for your support, encouragement, and tolerance; I learned a lot from you. Next I would like to thank people who I collaborated with: Axel, Thanks for helping me out with anything that relates to computers (!!), in particular for many Linux commands. Kostya, you are an ideal officemate. Thanks for all your help! I certainly can not count number of times you helped me along. I very much enjoyed the time we put on the motor proteins. Tatiana and Nienke, without you two I would have been really lonely on the coated-colloid topic. Thanks for sharing your ups and downs of your work! I would like to thank Sorin, Marileen, PieterRein, Bela, Benjamin, Dima, Angelo, and Pep for very useful discussions I had about my work.

Many people helped me with the preparation of this thesis: Rutger, Kostya, Simon, Sanne(h), Frank, Koos, Ana, Benjamin, Bianca, Nienke, Nima, Chantal, and Filipe. Thank you all for your suggestions, your critical reading, and your help with the design of the thesis. Without your help the thesis would not come out as nice as it is now. My thanks also goes to my Paranimfs, Maryam and Maryam for making the defense period more fun.

I would like to thank all the AMOLFers, my Iranian friends in the Netherlands, and many friends who are scattered around the world. I am very sorry, it feels really impossible to mention all your names. I have many nice memories of dinners, movies, parties, and trips that I had during the last four years. Thank you all for your support and your fantastic friendship. I would also like to thank Marija for her wonderful Alexander lessons.

Finally, I would like to thank my Mum, Batool, for her lovely support, and Shima and Nima, for their encouragement.

چکیده

در این کتاب، با استفاده از شبیه‌سازی رایانه‌ای به بررسی رفتار ذره‌های کولوییدی‌ای می‌پردازیم که می‌توانند به وسیله‌ی مولکول‌های زنجیری به‌هم متصل شوند. این زنجیرهای پلیمری که «تله‌کلیک» نامیده می‌شوند، با استفاده از دو انتهای چسبناک‌شان به صورت برگشت‌پذیر به سطح ذره‌های کولوییدی می‌چسبند. این انتهاهای چسبناک همچنین می‌توانند به انتهای یک زنجیر تله‌کلیک دیگر که ساختاری مکمل با زنجیر اولیه دارد، بچسبند و به این‌وسیله بین ذره‌های کولوییدی اتصال‌های پلیمری به وجود بیاورند. در فصل اول این رساله، مقدمه‌ی کلی‌ای درباره‌ی این مبحث ارائه شده است و بعضی از اصطلاح‌هایی که در این رساله از آنها بسیار استفاده میشود، را توضیح داده‌ام.

در فصل ۲، یک روش شبیه‌سازی مونته‌کارلوی گراند کانونی (GC) شبکه‌ای برای پلیمرها ارائه شده است. این روش به‌ویژه برای پلیمرهای تله‌کلیک که با سطح‌های موجود در سیستم برهمکنش دارند، مفید است. ترکیب چسبناکی قوی دو انتها و برهمکنش‌های ناشی از فضای اشغالی بین پلیمرها، باعث می‌شود روش‌های معمولی شبیه‌سازی کارایی خود را از دست بدهند. در این رساله، الگوریتمی برای این حالت پیشنهاد شده است که ترکیبی از روش انتشار تکانه Moment Propagation (MP) و روش پیکر بندی موثر Configurational Bias (CBMC) است. ایده‌ی اساسی این روش، استفاده از قدم‌زدن‌های کاتوره‌ای ناخودپرهیز برای ایجاد یک قدم‌زدن خودپرهیز روی شبکه است. الگوریتم پیشنهادی شرط تعادل جزییات (Detailed Balance) را ارضا می‌کند. در ادامه، کارایی این روش برای دو دسته مساله بررسی می‌شود: در بخش اول بر هم کنش تهینگی (Depletion) بین دو صفحه را حساب می‌کنیم. نتیجه‌ی به‌دست‌آمده با پژوهشهای قبلی به صورت کمی توافق دارد، هرچند در این حالت، روش MP در مقایسه با روش CBMC کارایی کمتری دارد. در بخش دوم برهمکنش دو صفحه در مخلوط زنجیرهای تله‌کلیک را بررسی می‌کنیم و نشان می‌دهیم برخلاف حالت قبل و به خصوص در حالتی که سطح‌ها بسیار نایک‌نواخت هستند، روشMP کارایی بیشتری نسبت به CBMC دارد. همچنین نشان می‌دهیم که کارایی نسبی MP با افزایش چگالی، افزایش پیدا می‌کند.

روشMP، در یک زمینه‌ی کاملا متفاوت بیوفیزیک نیز کاربرد دارد. از MP می‌توان به عنوان یک روش بدیع و بسیار حساس برای تشخیص پرش‌های یک موتور مولکولی در داده‌های پر اختلال حرکتش، استفاده کرد. حرکت کاینزین (Kinesin) بر روی میکروتیوبول‌ها (microtubules) نمونه‌ای از این حرکت است. نکته‌ی کلیدی این روش، که در بخش ۳ ارائه می‌شود، این است که معمولا کمیت مورد علاقه مسیر حرکت موتور نیست، بلکه اندازه‌ی پرش‌های موتور و تناوب این پرش‌ها است که اطلاعات لازم در مورد ساختار موتور را به ما می‌دهند. مسیرهای حرکت بسیاری ممکن است با داده‌های پر اختلال همخوانی داشته باشند، بنابراین بهتر است از تمام این مسیرها برای به دست آوردن اطلاعات مورد علاقه در مورد اندازه و تناوب پرش‌های یک موتور، استفاده شود. با استفاده از شباهت مسیر حرکت یک موتور مولکولی به یک گشت کاتوره‌ای جهت‌دار، الگوریتمی برای محاسبه‌ی «انرژی آزاد» این‌گونه گشتها با اندازه‌ی قدم‌های مختلف پیشنهاد می‌شود. این روش، به‌خصوص وقتی شدت اختلال داده‌ها بسیار زیاد است، کارایی بالاتری نسبت به روش‌های موجود دارد.

در بخش‌های ۴ و ۵ مدل درشت‌دانه‌ای ارائه می‌شود که برای شبیه‌سازی ذره‌های کولوییدی که با رشته‌های بلند DNAهای تله‌کلیکی پوشیده شده‌اند، مناسب است. انتهاهای چسبناک این رشته‌ها، رشته‌های کوتاه تک‌گانه و مکمل DNA هستند که می‌توانند به یکدیگر متصل شوند. در بخش ۴، روش مونته‌کارلوی گراند کانونی‌ای توضیح داده می‌شود که برای مطالعه‌ی خوداجتماع‌های این ذره‌ها به کار رفته است. در این روش با استفاده از تشابه یک رشته‌ی بلند DNA به یک کره‌ی نرم، برهمکنش میان کولوییدها بررسی می‌شود. در قسمت اول این بخش، از پژوهش‌های موجود و همچنین نتایج به دست آمده از شبیه‌سازی مونته‌کارلوی ارائه شده در این کتاب، که روی شبکه انجام شده است، استفاده کرده، به بررسی برهمکنش

یک پلیمر تله‌کلیک و یک کره‌ی کولوییدی می‌پردازیم. در قسمت دوم روشی را پیشنهاد می‌کنیم که چگونگی اتصال انتهاهای چسبناک را به ثابت‌های تعادلی دوگانه‌شدن دو رشته‌ی تک‌گانه‌ی DNA مربوط می‌کند.

در بخش ۵، از روش GCی ارائه شده در بخش قبل، برای شبیه‌سازی مخلوط معلق دو نوع ذره‌ی کولوییدی که با زنجیرهای DNA پوشانده شده‌اند، استفاده می‌کنیم. DNAهای چسبیده به این دو نوع ذره، مکمل هم هستند و می‌توانند به یکدیگر بچسبند. مقدار این دو ذره در مخلوط معلق مساوی است. در قسمت اول، شرایطی را در نظر می‌گیریم که انتهاها به شدت به یکدیگر می‌چسبند و در نتیجه هیچ انتهای آزادی در مخلوط معلق باقی نمی‌ماند. در این حالت، از حرکت‌های تعویضی برای تغییر آرایش پیوندهای زنجیری بین کولوییدها استفاده می‌کنیم. در این شرایط، یک گذار فاز مرتبه‌ی اول بین یک گاز از دوتایی‌های کولوییدی در چگالی پایین و یک مایع در چگالی بالا مشاهده می‌شود. این گذار فاز از افزایش آنتروپی پیوندهای زنجیری ناشی می‌شود. در قسمت دوم، به بررسی اثر مقدار «ثابت چسبنده گی » در رفتار سیستم می‌پردازیم. در این قسمت چسبندگی انتهاها چندان قوی نیست، در نتیجه بعضی از انتهاها آزاد می‌شوند. بنابراین از روش GC پیشنهاد شده در بخش ۴، به‌همراه حرکت‌هایی که پیوندها را قطع و وصل می‌کنند، استفاده می‌کنیم. نتایج به دست آمده نشان می‌دهند که گذار فاز مرتبه‌ی اول در دماهای پایین ولی غیرصفر (معادل چسبناکی محدود ولی غیربی‌نهایت)، نیز مشاهده می‌شود که به تدریج، بسته به تعداد زنجیرهای متصل به کولوییدها، در یک دمای بحرانی ناپدید می‌شوند.