Characterization of Nanopores With Internal Cavities for DNA Manipulation Using Langevin Dynamics Simulations

by

Martin Magill

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science

in

Modelling and Computational Science

University of Ontario Institute of Technology

Supervisors: Dr. Hendrick de Haan, Dr. Ed Waller

December 2016

Abstract

A novel nanopore geometry is proposed, in which a larger internal cavity is located inside a traditional nanopore. Polymer translocation through this geometry is studied using coarse-grained Langevin dynamics. The most striking result is that translocation time through the system is found to be minimal for polymers of medium length: both longer and shorter chains take longer to translocate. The length at which this occurs is named the critical length. This phenomenon arises as a balance between the driving electric force field and the entropic barrier that must be overcome in order for the polymer to exit the internal cavity. More detailed characterization of the system over a range of simulation parameters elucidate the physical mechanisms important to this mechanism. Using these results, a simplified free energy model is constructed and is solved analytically to predict the critical chain length as a function of applied field strength and cavity size. Good agreement is recovered between this theoretical model and numerical measurements over a range of parameters, and bounds of applicability are discussed. Applications of this new nanopore design are discussed.

Declaration

I, Martin Magill, declare that this thesis titled, Characterization of Nanopores With Internal Cavities for DNA Manipulation Using Langevin Dynamics Simulations and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Acknowledgements

I would like to thank Profs. Hendrick de Haan and Ed Waller, my supervisors, for guiding me through my Master's degree and making my thesis a success. Additionally, I would like to extend a special thanks to Prof. Hendrick de Haan, Prof. Ed Waller, and Dr. Dhavide Aruliah, for helping me out of a tight situation at the beginning of my studies. I would also like to thank Dr. Jeremy Pencer for sitting as the external reviewer on my defense committee. I'm also grateful to all the current and past members of the cNAB.LAB and the CLEAN group, as well as the faculty members of the Modelling and Computational Science program at UOIT. Further thanks go to my friends and family for the support given to me throughout my education. This work was possible thanks to the facilities provided by Compute/Calcul Canada, the Shared Hierarchical Academic Research Computing Network [1] and the cNAB.LAB.

Contents

Chapter 1

Introduction

Nanopores are nanometric holes in membranes. They are prevalent in nature, typically formed by protein stuctures passing through phospholipid bilayers. Such structures are referred to as biological nanopores. The α-hemolysin protein, which was one of the first nanopores studied in detail, is shown in Figure 1.1.

Nanopores can also be manufactured synthetically, most commonly by boring holes a few nanometers wide into membranes that are a few nanometers thick. The membranes are usually made of silicon nitride or similar materials borrowed from the manufacture of computer chips, and are therefore also called solid-state nanopores. Figure 1.2 illustrates the most basic possible structure for a synthetic nanopore: a cylindrical hole of constant radius. For the rest of this thesis, this style of nanopore will be referred to as a standard nanopore (SN). In this thesis, simulations are used to characterize a novel synthetic nanopore geometry, known as the cavity-nanopore (CN), illustrated in Figure 1.3. This work has also been accepted for publication in Physical Review Letters [46].

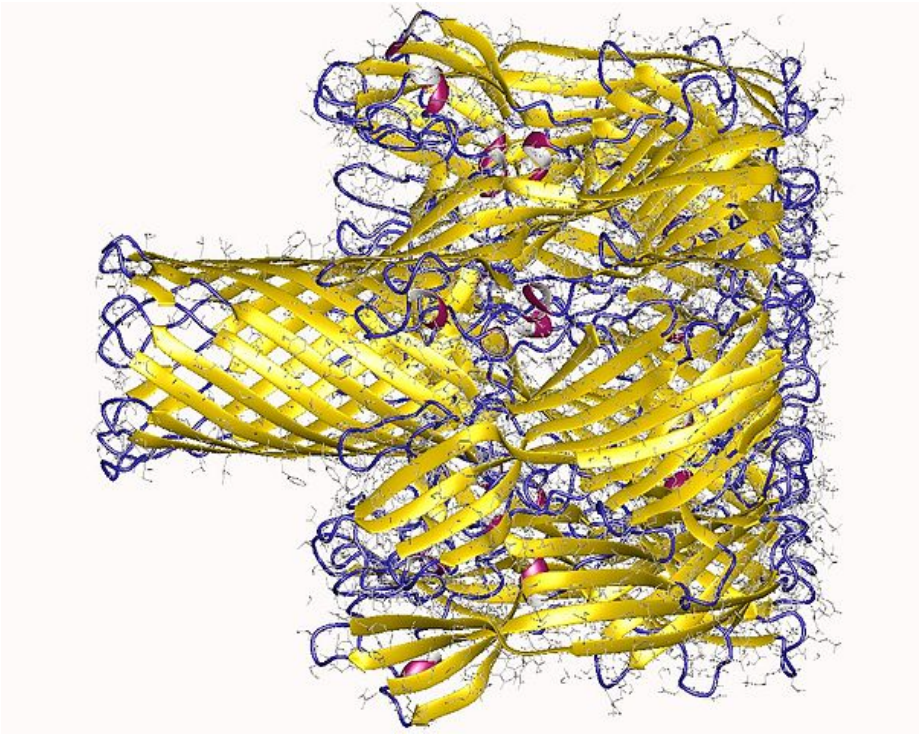

Figure 1.1: Visualization of alpha-hemolysin, the most well-studied biological nanopore [69].

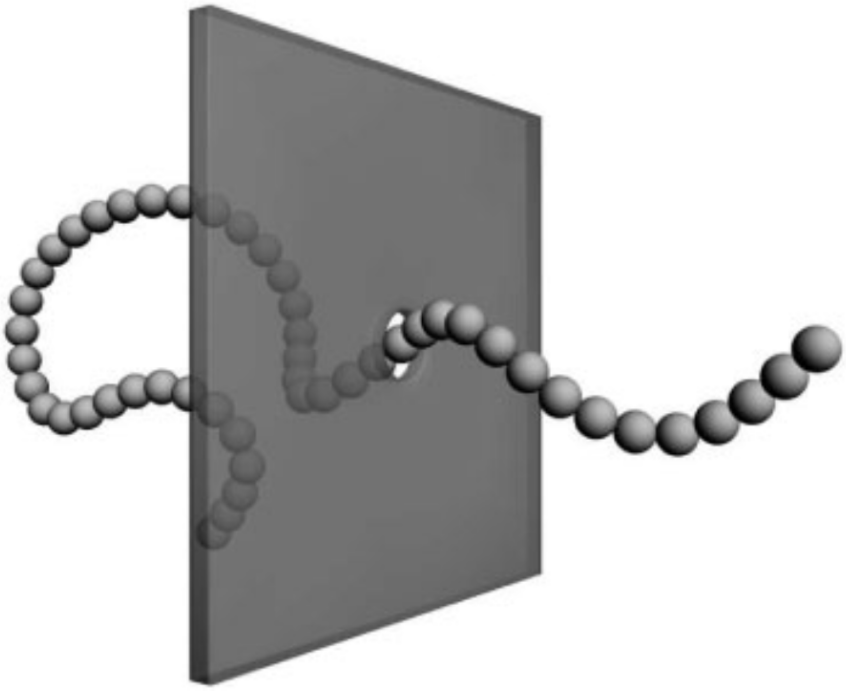

Figure 1.2: Illustration of a polymer translocating through a standard nanopore [8].

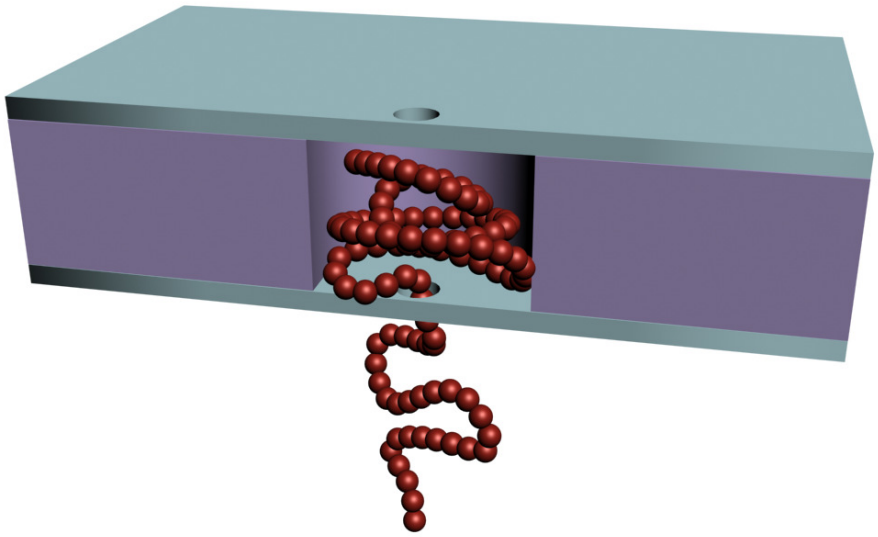

Figure 1.3: Illustration of a polymer translocating through the cavity-nanopore [46].

Both biological and synthetic nanopores are primarily of interest because of their interactions with polymers. Polymers are molecules formed by linking many similar sub-units together. The sub-units are referred to as monomers, and the process of forming polymers out of monomers is known as polymerization. Polymers can have various structures, and some of these are illustrated in Figure 1.4.

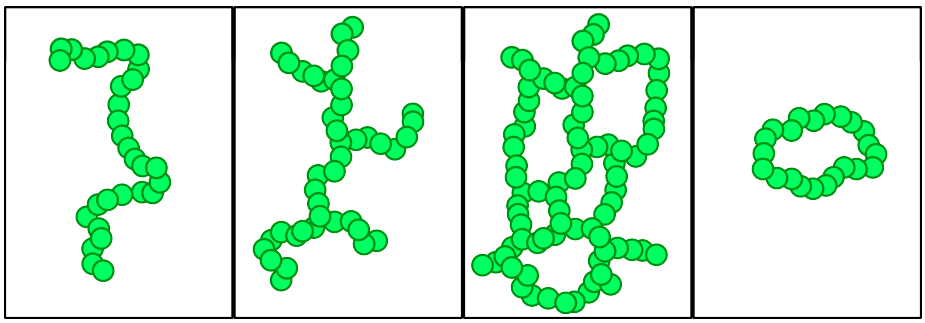

Figure 1.4: Example of polymer structures. From left to right, the illustrations show polymer that are linear, branched, cross-linked, and circular [73].

The simplest polymer structures, and those studied in this thesis, are linear polymers, where the monomers form a one-dimensional chain with two ends. Examples of such polymers include deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), which are fundamental to the storage and transport, respectively, of biological information in cells. Polymers occuring in biological systems are also referred to as biopolymers.

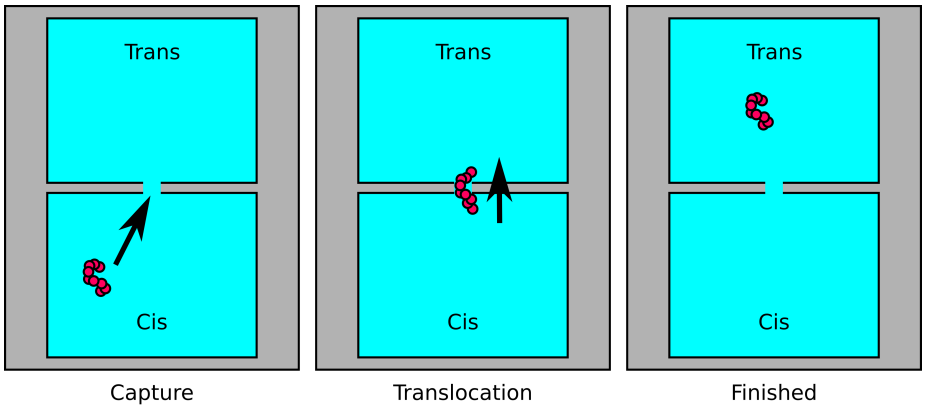

Figure 1.5: Schematic illustrating the stages of translocation. The polymer begins in the *cis* region, and diffuses randomly until it is captured by the pore. After capture, translocation begins. When translocation is complete, the polymer diffuses about the *trans* region.

The process of passing polymers through nanopores is known as translocation. The side on which the polymer begins is called the *cis* side, and the side to which it translocates is the *trans* side. This is illustrated in Figure 1.5. Translocation usually does not occur spontaneously, as it is very entropically unfavourable for a polymer to translocate rather than remain in free solution. It can be driven in a variety of ways (Figure 1.6): the polymer can be compressed on the *cis* side by a confining geometry, reducing the relative entropic barrier to translocation; the leading monomer of the polymer can be pulled through the nanopore e.g. by using optical tweezers; or a voltage difference can be applied across the membrane to drive electrically-charged polymers through the nanopore. The third option, using an external electric field, is easiest to implement experimentally, as many polymers of interest, including most biopolymers, are negatively charged. In biological systems, the voltage difference is created by the active transport of ions across the membrane. In synthetic nanopore experiments, the voltage difference is usually applied by electrodes.

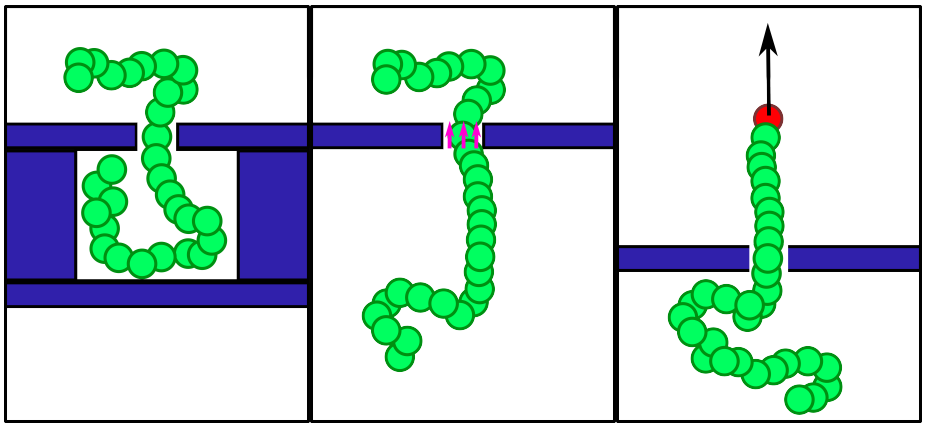

Figure 1.6: Schematics illustrating three common ways of driving polymer translocation. From left to right, they are driven by confinement, an electric field across the pore, and an optical tweezer pulling the leading monomer.

When a voltage is created across a membrane with a nanopore in it, any ions present in the electrolyte will flow in response to the electric field, creating an ionic current through the system. When a polymer translocates through the nanopore, it reduces the cross-section of the nanopore through which the ions can flow, creating a concomitant drop in the ionic current[1].

In nanopore experiments, the ionic current is sampled at high frequencies so as to resolve the detailed variations in current occuring during individual translocation events. Such data is called a current trace, and an example is shown in Figure 1.7. If the translocating polymers are DNA molecules, the magnitude of the current drop at any point in time corresponds to the type of nucleotide base(s) present in the nanopore at that time. As a result, the current trace can be used to infer the sequence of the DNA molecule. DNA sequencing technology is highly sought-after, motivating much of the research

[1]This description of the current blockade effect is actually simplified. Detailed atomistic simulations by Kesselheim et al. suggest that a translocating DNA molecule actually increases the concentration of ions near the pore, which should facilitate ion exchange through the hole and increase current [35]. Indeed, an increase in current during translocation can be observed under appropriate experimental conditions, e.g. at low salt concentrations [68]. The decrease in current typically observed may be due to a molecular drag felt by conducting ions near the surface of the DNA molecule [35].

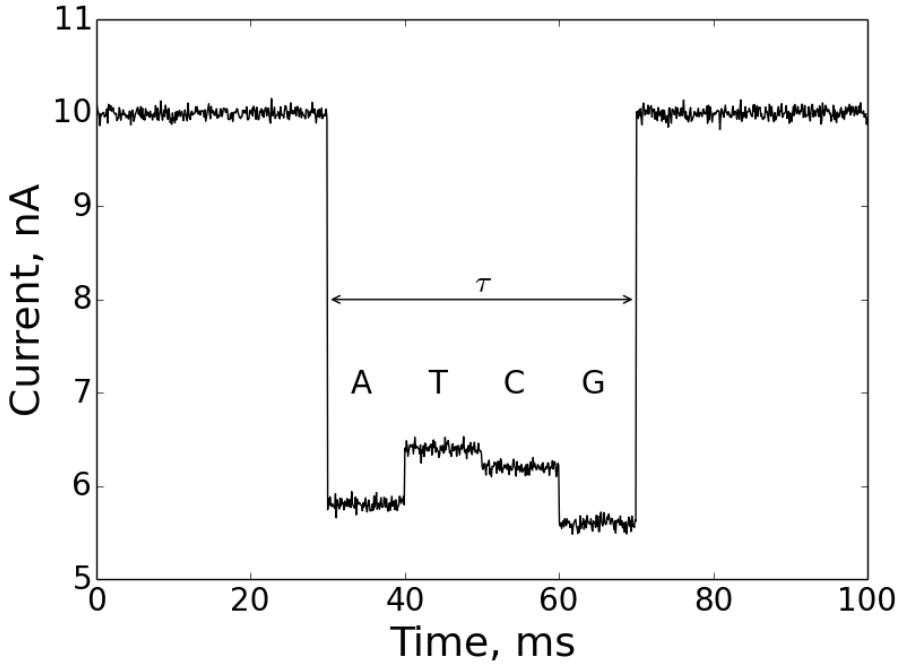

Figure 1.7: Example of a current trace of a hypothetical DNA translocation event (not real data). The translocation time is denoted τ. The variations in the current during the event can be mapped to the base pair in the nanopore at that point in time.

and development that has gone into nanopores.

In addition to sequencing applications, the current trace can be used to determine the duration of translocation events, which contains information about the length of the polymer. The event begins when the current drops significantly below its baseline value, and continues until the current returns to this baseline, yielding an experimental measurement of the translocation time, denoted τ. Figure 1.7 illustrates the measurement of τ from a current trace. The data in Figure 1.7 is completely fabricated, and serves only to illustrate the concept of a current trace schematically.

The average τ is a function of the polymer length. If this relationship is

known, measurements of τ can be used to estimate polymer length. Such length measurements can usually be made more rapidly and more reliably than DNA sequencing measurements. Furthermore, all polymers can be analysed in this fashion, whereas sequencing and other detailed chemical analyses are specific to certain types of polymers.

The study of translocation times as a function of polymer length has been the focus of hundreds of experimental, theoretical, and simulation-based studies. In general, for a membrane with a standard nanopore, all studies agree that longer polymers take longer to translocate on average, i.e. the average translocation time $\langle\tau\rangle$ is a monotonic function of chain length. However, the exact relationship between polymer length and translocation time is complex and subtle. This relationship will be explored in more detail in the literature review.

1.1 Nanopore Applications

As RNA and DNA are the fundamental units of information in all biological systems, the use of nanopores to detect, manipulate, and read these molecules has inspired applications in many fields.

1.1.1 Rapid DNA Sequencing

Oxford Nanopore Technologies is currently the leader in commercial nanopore devices. Their MinION device contains up to 512 nanopores that sequence single-stranded DNA in parallel. DNA sequencing has a plethora of applications, including but not limited to the following:

- **Forensic Science:** DNA is used to place persons of interests at the scene of a crime since Sir Alec Jeffreys pioneered DNA fingerprinting in 1984 [57].

- **Long-Term Data Storage:** Recent start-ups like Twist Bioscience are interested in using synthetic DNA molecules to store information in a format that is stable on long time scales [4]. Microsoft recently purchased 10 million synthetic strands for this purpose [4].

- **Pathogen Identification:** Bacteria and viruses can be uniquely identified by their genetic sequences. Samples of interest can be analysed for any genetic material that might indicate the presence of lifeforms of interest. For instance, Cao et al. used a DNA microarray technique to analyse samples of fishery products [5].

- **Personalized Disease Forecast:** The genomic information encoded in DNA can be used to predict an individual's predisposition to certain diseases [34]. Companies like 23andMe have marketed commercial human genome sequencing, so that private individuals can pay to have their genes analysed for such markers.

- **Cancer Diagnosis**: The detection of certain RNA and DNA molecules in blood samples may be useful in screening for and diagnosing cancers [32].

1.1.2 Sorting Polymers by Length

Despite the promises of rapid DNA sequencing, it remains unattained as a technological feat. In particular, as of 2015 the MinION nanopore-based rapid DNA sequencing device marketed by Oxford Nanopore Technologies was estimated to have an error rate of roughly 38.2% in practice [39]. This large error rate limits its ability to compete with existing sequencing technologies [39]. Nonetheless, nanopores are also being considered for use in many other promising technological applications.

Most current laboratory techniques involving DNA analysis do not attempt to explicitly sequence DNA molecules in their entirety. Instead, solutions of mixed DNA molecules are sorted by their lengths, and the spectrum of DNA fragment sizes is analysed. Traditionally, the method employed for this sorting process is one called gel electrophoresis.

In gel electrophoresis, the DNA solution is placed in a gel across which a uniform electric field is applied. As DNA molecules carry a negative charge, they are forced through the gel towards the positive terminal of the applied electric field. As discussed later, DNA molecules are free-draining, so that they all move with equal drift velocities when placed in a liquid solution and subjected to an electric field. Gels, however, are composed of cross-linked molecular networks dissolved within a liquid. As the DNA molecules are forced through these networks by the electric field, they are impeded by the cross-linking chains. Long DNA chains become entangled in these networks, whereas smaller DNA fragments are significantly less encumbered. The net result is that smaller chains move through the gel more rapidly, on average, than longer chains: the mobility of a DNA molecule forced through a gel is a decreasing function of its length.

If a mix of DNA molecules is placed closely together at one end of the gel and subjected to this process, the sample will become separated according to DNA size. Unfortunately, this process takes from several hours to several days to achieve adequate separation, and requires a significant amount of manual labour. These challenges have prompted the development of many new electrophoretic separation techniques.

Nanopores show great promise as a tool for DNA separation. In gel electrophoresis, DNA molecule mobility becomes a function of chain length due to interactions with the gel network. Similarly, translocation time through a nanopore is a strong function of chain length, as discussed previously. Forc-

ing DNA chains through a series of nanopores would therefore have a similar separation effect to that exploited in gel electrophoresis.

In contrast to gel electrophoresis, nanopore-based separation would have several advantages. Whereas gel electrophoresis takes hours or days, nanopore-based separation should occur far more rapidly [67]. Furthermore, gel electrophoresis requires manual labour and access to a laboratory, whereas nanopore-based devices are small enough to be incorporated into lab-on-a-chip designs.

Finally, using novel nanopore geometries (like that presented in this thesis), the relationship between chain length and mobility through the nanopore-based separation device could be tailored to affect the order into which the chains will be sorted. Indeed, the primary application motivating the design of the cavity-nanopore is that of sorting polymers by length. This feature will be discussed again in the Conclusion of the thesis.

1.1.3 Other Nanopore Applications

As DNA sequencing has such an enormous range of applications, it has been the primary driver for nanopore research. Sorting DNA by length is another major area of research, and the primary motivation for the work in this thesis. Additionally, many alternative applications for nanopores have been proposed and explored, including but not limited to the following [49].

- Detect, identify, and count analytes in small concentrations [18, 3]
- Mass spectrometry for polymers in solution [56]
- Identify drug stereoisoforms [29]
- Sort proteins by length [70, 77]
- Detect microRNA molecules [19]

Nanopores have also been used to manipulate DNA and RNA molecules in order to study various biological processes at the nanoscale, such as the following [54].

- Unzipping of double-stranded DNA [23, 64, 11, 74]
- DNA-protein interactions [22, 16]
- Helix-coil transitions [42]
- The DNA replication process [7, 40]

Despite the many possible alternative applications of nanopores, DNA sequencing remains the area of primary focus [49].

1.2 Introduction to Polymer Physics

Before exploring the existing nanopore literature, it is necessary to review some aspects of basic polymer physics. In this section, important concepts and terminology will be reviewed. In the next section, polymer models that can be simulated in a computer will be discussed.

As stated previously, a polymer is any molecule formed by linking together many copies of simpler units. These building blocks are referred to as monomers. Polymers can be made up of various monomer types: DNA, for instance, is made up of four nucleotide bases. Polymers that only contain one monomer type are known as homopolymers: polyvinyl chloride (PVC), for instance, is formed only of vinyl chloride monomers. Polymers can also have various structures, as illustrated earlier in Figure 1.4. This thesis will focus on linear polymers.

Consider the polymer shown in Figure 1.8. This model of a generic linear homopolymer will be the basis for the discussion in this section. The chain is made of N identical spherical monomers, and neighbouring monomers are

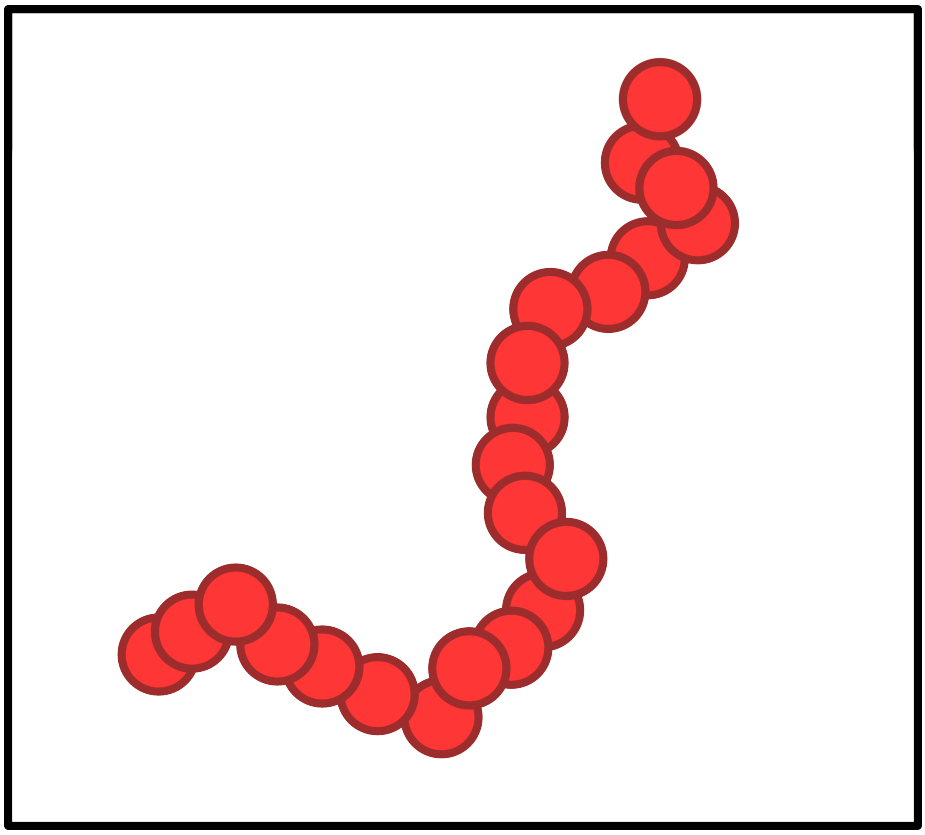

Figure 1.8: A model of generic linear homopolymer. The spheres are monomers, which are bonded together along the polymer.

bonded together by springs. The average center-to-center distance between bonded monomers is called the bond length, l. The whole polymer exists in a bath of solvent particles. The monomers are constantly jostled by interactions with the solvent, so that the shape of the polymer is always changing.

1.2.1 Describing Polymer Conformations

The shape of a polymer at any instant in time is known as its conformation. The average conformation of a polymer over time and the rate at which a polymer's conformation changes are important questions in polymer physics. This

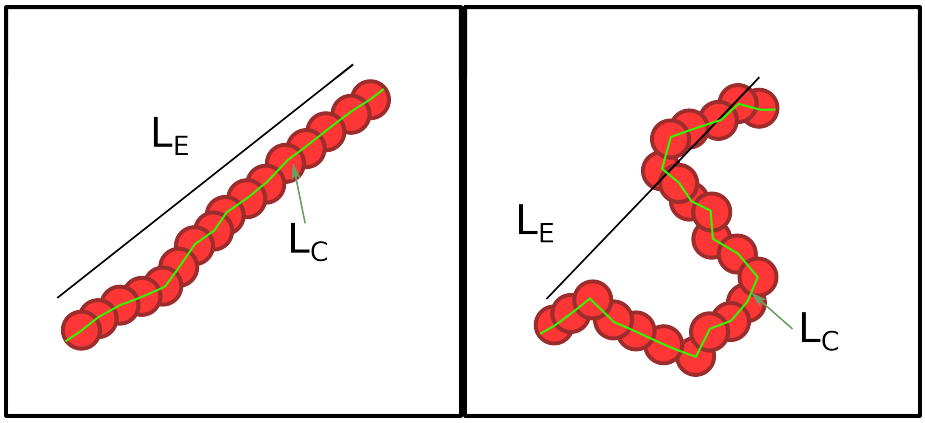

Figure 1.9: Illustration of contour length L_C and end-to-end length L_E on rod-like (left) and flexible (right) polymers.

section will focus on describing a polymer's size, which is but one aspect of the total conformation. However, as polymers are dynamic structures without well-defined boundaries, measuring their size is non-trivial.

The simplest way to estimate a polymer's size is by its length, measured from one end to the other along the centers of the monomers. This length is known as the contour length L_C of the polymer, and its average value is $\langle L_C \rangle = (N-1)l$ [73]. Countour length is a useful measure of size for stiff, rod-like polymers. However, flexible polymers are more often in coiled conformations. Figure 1.9 illustrates the difference. Clearly, countour length is not as good a measure of dimensional extent for flexible polymers.

Two other metrics of polymer size are more useful in conveying the volume occupied by a flexible polymer. The first is the linear distance measured between the two free ends of the polymer, known as the end-to-end distance, L_E. The end-to-end distance is compared to the contour length in Figure 1.9. The other measure of size is called the radius of gyration, R_G, which is a measure of how much the polymer is spread.

The radius of gyration can be defined as the root-mean-square distance be-

tween all pairs of monomers, that is

$$R_G^2 = \frac{1}{2N^2} \sum_{i,j} (\mathbf{r}_i - \mathbf{r}_j)^2 . \tag{1.1}$$

Each term $(\mathbf{r}_i - \mathbf{r}_j)^2$ is the square distance between monomer i and monomer j. Figure 1.10 illustrates the radius of gyration, and compares it to L_C and L_E.

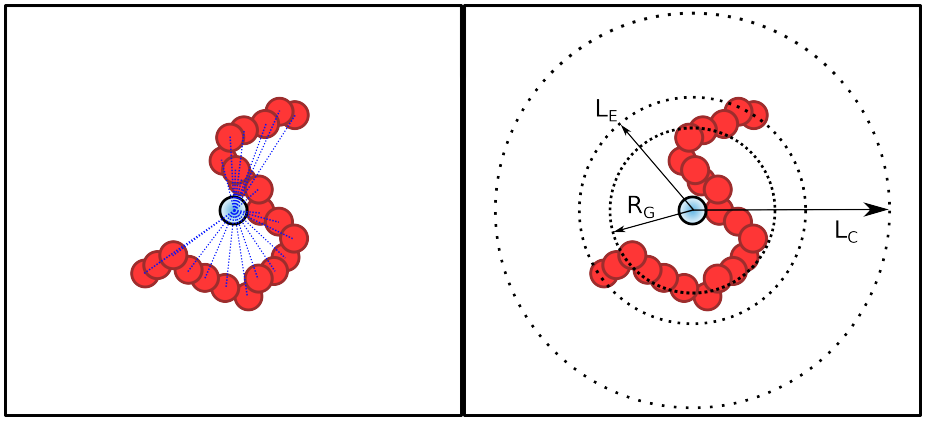

Figure 1.10: The left schematic illustrates the calculation of the radius of gyration. The shaded monomer indicates the average monomer position, so R_G^2 is the average square length of the blue dotted lines. The right schematic compares radius of gyration R_G, contour length L_C and end-to-end length L_E.

The scaling factor in the formula for R_G makes Equation 1.1 equivalent to the standard deviation of monomer positions:

$$\begin{aligned} R_G^2 &= \mathrm{Var}_i(\mathbf{r}_i) && (1.2)\\ &= \left\langle (\mathbf{r}_i - \langle \mathbf{r} \rangle_N)^2 \right\rangle_N && (1.3)\\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{r}_i - \frac{1}{N} \sum_{i=1}^{N} \mathbf{r}_i \right)^2 , && (1.4) \end{aligned}$$

where $\langle \cdot \rangle_N$ is used to indicate an average that is taken over the N monomers, as opposed to an average that is taken over time [73]. This in turn can be written

as

$$R_G^2 = \left\langle \left(\mathbf{r}_i - \langle \mathbf{r} \rangle_N\right)^2 \right\rangle_N \tag{1.5}$$

$$= \left\langle \mathbf{r}^2 \right\rangle_N - \langle \mathbf{r} \rangle_N^2 . \tag{1.6}$$

Equation 1.6 is the most computationally efficient form for computing R_G, as it can be computed using a single `for` loop [73]. Conversely, the other forms require two `for` loops or nested `for` loops. Figure 1.10 illustrates R_G calculated using Equation 1.6.

Theoretical formulae for L_E and R_G can be calculated in some cases. First, consider a polymer where monomers only interact with their neighbours, i.e. the monomers to which they are bonded. In other words, monomers that are not bonded to one another can pass right through one another. All bonds are oriented independently of all other bonds. Such a polymer is called an ideal chain or a freely-jointed chain. In this case, the conformation of the polymer at any instant in time is simply a random walk of average step size l. The average end-to-end distance L_E of an ideal chain is equal to the average displacement in a random walk, which is zero:

$$\langle L_E \rangle_t = 0, \tag{1.7}$$

where $\langle \cdot \rangle_t$ is now used to indicate an average that is taken over time, as opposed to $\langle \cdot \rangle_N$ [73]. The standard deviation of L_E over time is equal to the standard deviation of the displacement of a random walk,

$$\left\langle \left(L_E(t) - \langle L_E \rangle_t\right)^2 \right\rangle_t = \left\langle \left(L_E(t)\right)^2 \right\rangle_t = l\sqrt{N} \quad [73]. \tag{1.8}$$

The radius of gyration of an ideal chain is

$$R_G = l\sqrt{\frac{N}{6}} \quad [73]. \tag{1.9}$$

Thus, both of these metrics of polymer size scale as $\sqrt{N}$ for an ideal chain.

Now consider a polymer where monomers cannot overlap. Specifically, let the monomers be hard spheres of diameter σ, such that no two monomers can have a center-to-center distance less than σ. Monomers that are farther than σ apart do not interact unless they are neighbours. Such an interaction is called excluded volume, and this polymer model will be referred to as a freely-jointed chain with excluded volume.

Adding excluded volume interactions to the polymer model makes it larger than the ideal chain. This is intuitive: monomers in the ideal chain were capable of occupying the same volume simultaneously, so an ideal chain can be packed into a smaller volume than a chain with excluded volume. In this chain, the radius of gyration scales as

$$R_G \propto N^{\nu}, \tag{1.10}$$

where $\nu \approx 0.588$ is called the Flory exponent [73]. The end-to-end distance L_E scales the same way [73].

The coefficient of proportionality between R_G and N^{ν} is difficult to obtain theoretically and depends on the specific implementation of the excluded volume. In practice, the appropriate coefficient for a given implementation can be measured from simulations. Figure 1.11 shows such measurements for the simulation technique used in the next chapter. From this, the coefficient of proportionality can be estimated as 0.402. Incidentally, this is close in value to the coefficient of proportionality in the ideal chain case, $1/\sqrt{6} \approx 0.408$.

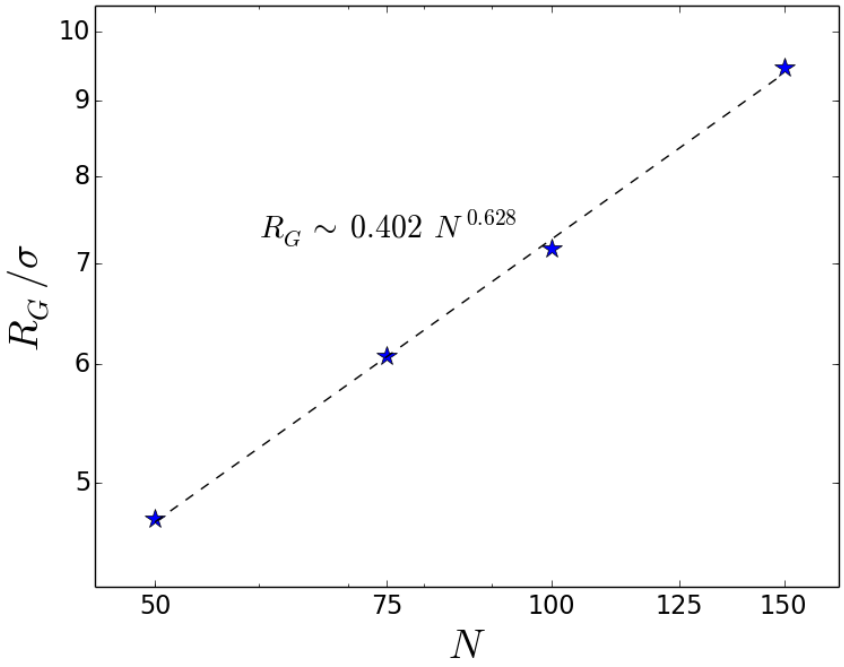

Figure 1.11: Plot of measured R_G values from simulation as a function of chain length N. The power law of best fit is $R_G = 0.402N^{0.628}$.

In summary, the introduction of excluded volume interactions to the polymer model means that both size metrics now scale as N^ν, where $\nu \approx 0.588 > 0.5$. This is in agreement with the expected result, which is that excluded volume interactions should lead to larger polymers.

1.2.2 Polymer Relaxation

The previous section establishes the terminology required to describe polymer conformations, and gives formulae for the sizes of ideal chains and freely-jointed chains with excluded volume. This section will address the evolution of polymer conformations over time.

Consider a polymer in a specific conformation C_1 at a time t_1. At a second time t_2 that occurs very shortly after t_1, the polymer will be in a conformation

C_2. If $t_2 - t_1$ is sufficiently small, the monomers will only have moved a small distance. As such, C_2 will be very similar to C_1. Conversely, consider the polymer's conformation C_3 at a time t_3 that is a long time after t_1. The polymer has been subjected to many random fluctuations in the time between t_3 and t_1, and so we expect its final conformation C_3 to be almost completely unrelated to its initial conformation C_1. That is, the polymer's conformation changes over time, gradually loses any correlation to its earlier conformations.

The timescale on which a polymer's conformation changes appreciably is given by its relaxation time, denoted τ_R. Another way of describing τ_R is that it is the timescale over which the polymer's final conformation becomes uncorrelated with its initial conformation. Clearly the exact definition of relaxation time depends on how one measures the correlation between two conformations. For this study, we will use the end-to-end distance L_E as a proxy for the polymer's evolution in time. It behaves as desired: over short timescales, L_E will only change slightly; over longer timescales, L_E will be subject to many random fluctuations and, on average, its final value will be unrelated to its initial value.

We will quantify the correlation between the end-to-end distance $L_E(t)$ at some time t and its value some time later, $L_E(t + \Delta t)$, by

$$\mathrm{Corr}_{L_E}(t, t + \Delta t) = L_E(t)L_E(t + \Delta t). \tag{1.11}$$

When Δt is small, the two factors are approximately the same, so $\mathrm{Corr}_{L_E}(t, t + \Delta t) \approx L_E(t)^2$.

Conversely, if Δt is large, $L_E(t)$ and $L_E(t+\Delta t)$ will be independent measurements of the end-to-end distance of the same polymer. The value of $\mathrm{Corr}_{L_E}(t, t + \Delta t)$ at any given t will be random, but its expected value can be computed. In particular, since the measurements $L_E(t)$ and $L_E(t + \Delta t)$ are essentially independent for large Δt, the expected value of their product $L_E(t)L_E(t + \Delta t)$

approaches the product of their expected values. Furthermore, their expected values are each $\langle L_E \rangle_t = 0$ as discussed in the previous section. In other words,

$$\lim_{\Delta t \to \infty} \langle \mathrm{Corr}_{L_E}(t, t+\Delta t) \rangle_t = \lim_{\Delta t \to \infty} \langle L_E(t) L_E(t+\Delta t) \rangle_t \tag{1.12}$$

$$= \langle L_E \rangle_t^2 = 0. \tag{1.13}$$

This precise definition of correlation can now be used to derive the timescale over which correlation decays. Consider the autocorrelation function defined by

$$\mathrm{AutoCorr}_{L_E}(\Delta t) = \langle \mathrm{Corr}_{L_E}(t, t+\Delta t) \rangle_t \tag{1.14}$$

$$= \langle L_E(t) L_E(t+\Delta t) \rangle_t . \tag{1.15}$$

That is, the autocorrelation function of L_E for a given Δt is the expected value of the correlation between L_E at some arbitrary initial time and L_E some time Δt after. Informally, $\mathrm{AutoCorr}_{L_E}(\Delta t)$ quantifies how much L_E changes over a timescale Δt.

Since $\mathrm{Corr}_{L_E}(t,t) = L_E(t)^2$, we have that $\mathrm{AutoCorr}_{L_E}(0) = \langle L_E^2 \rangle$, for which formulae were given in the previous section. Regardless of the polymer model, this will be some finite positive number. Conversely, we just showed that $\mathrm{Corr}(t, t+\infty) \to 0$, so $\mathrm{AutoCorr}_{L_E}(\infty) \to 0$. Thus the autocorrelation function has a positive value on short timescales and goes to zero on very long timescales, in accordance with the informal description of polymer relaxation given above. In particular, it turns out that the autocorrelation function has the form

$$\mathrm{AutoCorr}_{L_E}(\Delta t) = \langle L_E^2 \rangle \exp\left(-\frac{\Delta t}{\tau_R}\right) \quad [73]. \tag{1.16}$$

That is, the autocorrelation function decays exponentially [73]. For this thesis,

we will define the relaxation time to be precisely this timescale τ_R on which AutoCorr_{L_E} decays. This provides us with a rigorous definition of relaxation time.

Other relaxation times can be defined from the autocorrelation functions of other metrics related to polymer conformation, such as R_G. Conversely, theoretical estimates of relaxation time can be made by considering polymer dynamics. These will be introduced in the next section, after polymer dynamics are discussed.

1.3 Introduction to Simulations for Polymer Physics

The previous sections introduced terminology and metrics used to describe the shape and size of polymers and the rate at which these quantities evolve in time under the influence of random fluctuations. These concepts were illustrated using two polymer models: the ideal chain and the freely-jointed chain with excluded volume.

This section will introduce techniques for implementing numerical simulations of polymers. In particular, the freely-jointed chain with excluded volume will be implemented in the coarse-grained Langevin dynamics framework.

1.3.1 Molecular Dynamics

In molecular dynamics (MD), physical systems are represented by N particles, and the system is evolved using Newton's second law,

$$m_i \mathbf{a}_i(t) = \mathbf{F}_i(t), \tag{1.17}$$

where m_i is the mass of the ith particle in the system, $\mathbf{a}_i(t)$ is its acceleration at time t, and $\mathbf{F}_i(t)$ is the net force acting on that particle at time t. The

forces acting on the ith particle are due to interactions with all remaining $N-1$ particles.

MD can be used to model polymer systems at the atomic scale. The polymer, the solvent, and any material surfaces (such as a silicon nitride membrane) are constructed out of atoms according to their molecular structures. The space between the N particles is filled with vacuum. Forces in such a system are usually electrostatic in nature.

Although these atomistic MD simulations allow polymers and solvent to be represented in astounding detail, they have the disadvantage of being very computationally expensive. Most polymers of interest, such as DNA, contain an enormous number of atoms, and simulations must also include the dense solvent environment surrounding the polymer. In particular, it isn't computationally feasible to conduct the simulations in this thesis in atomistic detail.

1.3.2 Coarse-Grained Langevin Dynamics

Instead of attempting to represent polymer systems in complete physical detail, this thesis will construct models of the system that are sufficiently sophisticated to capture complex polymer phenomena, while remaining simple enough to be simulated on large length and long time scales using available modern computers. The simulation will use coarse-grained polymers with an implicit solvent. The nanopore membrane will be represented by an internal boundary condition. The following sections explain these methods in detail.

In coarse-grained langevin dynamics (CGLD), each monomer of the polymer will be represented by a single particle. In this way, the model can be used to implement the generic polymer models from the previous section directly. This section will begin by introducing the equations of motion for individual monomers when they are not part of a polymer. Then, the freely-jointed chain

with excluded volume will be implemented inside this formalism. Finally, the internal boundary condition used to implement the nanopore membrane will be explained.

Consider a single particle, representing a monomer, that exists in a large bath of solvent. In MD, the interactions of the monomer with the solvent were calculated from the detailed atomic composition of the monomer and the solvent. This was found to be very computationally expensive, in part because of the very large number of solvent molecules.

Conversely, Langevin dynamics (LD) takes advantage of this fact: if the monomer is moving in a large population of solvent molecules, and the dynamics of the solvent are not of primary interest, then the solvent can be approximated as a continuous fluid. On the lengthscales of interest for nanopore systems, the solvent model must capture two important interactions with the monomers. First, the solvent must exert a drag on the monomers. Second, the solvent must impart random thermal fluctuations to the motion of the monomers. The LD equations of motion will be created starting from the MD equations of motion for monomers in a vacuum, to which two forces are added to represent these two solvent interactions.

1.3.2.1 Drag in a Viscous Fluid

The drag experienced by objects moving in a fluid has been the subject of enormous amounts of study. Drag behaves very differently for objects that move quickly with respect to the fluid than for those that move slowly. This can be quantified by the Reynolds number, Re, defined as

$$\mathrm{Re} = \frac{vL}{\mu/\rho}, \tag{1.18}$$

where v is the characteristic speed of the particle relative to the fluid, L is the characteristic length scale of the object, μ is the dynamic viscosity of the fluid, and ρ is the density of the fluid. Viscosity is a measure of the intrinsic thickness of a fluid. High Reynolds numbers correspond to thin fluids, whereas low Reynolds numbers correspond to thick fluids. The distinction between the two types of fluids is usually around $\mathrm{Re} \approx 10^3$.

The Reynolds number indicates that a fluid that feels thin under certain conditions may feel thick under different conditions. For instance, water has $\mu/\rho \approx 10^{-6}$ Pa·s, in SI units. For a human hand moving in a bathtub, $v \approx 0.1$ m/s and $L \approx 0.1$ m, so $\mathrm{Re} \approx 10^4$. This is a high Reynolds number, so water behaves as a thin fluid. Conversely, for a DNA strand moving near a nanopore in water, $v \approx 10^{-6}$ m/s (e.g. translocation speed of 3 bp/ms [14]) and $L \approx 10^{-9}$ m, so $\mathrm{Re} \approx 10^{-9}$. This is an extremely low Reynolds number, so water behaves as a very thick fluid in nanopore systems.

It is known that for fluid systems of very low Reynolds number ($\mathrm{Re} < 1$), the drag experienced by a spherical object moving through the fluid obeys Stokes' law:

$$\mathbf{F}_{\mathrm{drag}} = -6\pi\mu r_s \mathbf{v}, \tag{1.19}$$

where r_s is the radius of the sphere and $\mathbf{v}$ is its velocity relative to the fluid [2]. The quantity $\gamma = 6\pi\mu r_s$ is called the friction coefficient, so that the force can be rewritten more simply as

$$\mathbf{F}_{\mathrm{drag}} = -\gamma \mathbf{v}, \tag{1.20}$$

This will be the first force added to the MD equations to create the LD equations of motion for the monomers.

1.3.2.2 Random Forcing Term

The second interaction between the solvent and the monomer arises from the particle nature of the solvent, and cannot be derived from the fluid perspective of the solvent. Consider again the MD formulation, but now suppose that the monomer is at rest with respect to the average motion of the solvent molecules, so that it does not experience the drag force discussed above. Nonetheless, since the solvent is a thermal system, the solvent molecules are constantly in thermal motion and will collide randomly with the monomer.

These random collisions with solvent molecules jostle the monomer by small amounts at a time. Since the monomer is at thermal equilibrium with the solvent molecules, any individual collision will not impart much kinetic energy to it. However, sometimes many collisions with different solvent molecules will happen at the same time and in the same direction. Alternatively, a particularly energetic solvent molecule may collide with the monomer. Both situations create non-negligible forces on the monomer.

The net effect of these random collisions is represented in Langevin dynamics by a random force term. The term is proportional to a so-called stationary Gaussion process with zero mean, $\mathbf{R}(t) = (R_x(t), R_y(t), R_z(t))$, that satisfies

$$\langle R_i(t) \rangle_t = 0, \tag{1.21}$$

$$\langle R_i(t_1) R_i(t_2) \rangle_t = \delta(t_2 - t_1) \tag{1.22}$$

for $i = x, y, z$. In other words, the time-average of this random force term is zero, and the force at any given moment in time acts independently of the force at all other times. Furthermore, the three components of $\mathbf{R}(t)$ are mutually

independent. Thus, the random force term is of the form

$$\mathbf{F}_{\text{random}} = K\mathbf{R}(t). \tag{1.23}$$

The magnitude K of the random force term is a function of the thermal properties of the solvent. The temperature T of the system controls the frequency with which the solvent collisions occur: a hotter solvent fluctuates more rapidly, so the monomer will experience more collisions per unit time. However, these thermal properties are also coupled to the viscosity of the fluid. The relationship between the temperature of the solvent and the friction coefficient experienced by a given monomer is given by the fluctuation-dissipation relation,

$$D = \frac{k_B T}{\gamma}. \tag{1.24}$$

Here, k_B is Boltzmann's constant and D is a constant known as the diffusion coefficient.

From this coupling between the drag force and the thermal fluctuation force, the magnitude of the latter can be found to be

$$\mathbf{F}_{\text{random}} = \sqrt{2k_B T\gamma}\mathbf{R}(t) \quad [65]. \tag{1.25}$$

This is the second force term that defines solvent interactions in the LD system. The LD solvent model is therefore complete.

1.3.2.3 Langevin Dynamics

The LD equation of motion for a single monomer alone in a solvent is thus

$$m\mathbf{a} = -\gamma\mathbf{v} + \sqrt{2k_B T\gamma}\mathbf{R}(t). \tag{1.26}$$

This is known as the second-order Langevin equation. From this, a polymer can be formed by introducing other monomers and defining appropriate force terms between the monomers. This will be covered in the next section. Before that, however, it is useful to consider the dynamics of a single free monomer acting under this equation of motion.

1.3.2.3.1 Diffusion of a Free Monomer The motion of a particle in LD in the absence of other particles and external force fields is called Brownian motion. The particle moves around randomly, driven by the random forcing term. Any inertia the particle accumulates from this forcing is rapidly damped out by the drag force. This is called the overdamped limit. Brownian motion has some very important universal properties.

Since Brownian motion is a random process, it is best described by statistical quantities. In particular, it is easiest to discuss the behaviour of a large population of identical, non-interacting particles each undergoing independent Brownian motion. In this picture, the individual motion of the particles causes the population to spread out over time. This process is called diffusion, and it represents the average behaviour of particles undergoing Brownian motion.

Suppose very many particles start at the origin and begin diffusing. We are interested in quantifying how much diffusion is occuring. If the particles move in the same fashion in all three spatial dimensions, then by symmetry the average position of the population of particles must remain at the origin as diffusion proceeds. Thus the average position is not a good measure of diffusion.

Since diffusion is a spreading process, it is better quantified by the variance of the particle positions. The variance is the average square distance from the average particle position. If the particles are at positions $\mathbf{r}_i$ then this is

$$\mathrm{Var}(\mathbf{r}_i) = \left\langle (\mathbf{r}_i - \langle \mathbf{r} \rangle)^2 \right\rangle . \tag{1.27}$$

In the limit of many particles and long timescales, the variance of a population of particles undergoing Brownian motion in three dimensions satisfies

$$\left\langle \left(\mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle\right)^2 \right\rangle = 6Dt, \tag{1.28}$$

where diffusion has been occuring for a duration of time t. The coefficient D is the same diffusion coefficient introduced in Equation 1.24. Thus, all diffusive Brownian motion leads to a spreading out of the population that grows linearly in time. The specific nature of the diffusing particles is captured in the diffusion coefficient.

The previous discussion considered a population of identical, non-interacting particles, but the results also apply to the average behaviour of a single diffusing particle. As before, if the particle moves identically in all three spatial dimensions then its average position over time must be its initial position by symmetry. In the case of the population, this led to the use of the variance in particle positions to quantity how much diffusion has occured. In the case of one particle, the variance of the population is replaced by a quantity called the mean square displacement. It is the expected value of the square distance between a particle's initial and final positions. If the particle is at position $\mathbf{r}(t)$ at time t, then its mean square displacement is

$$\left\langle \left(\mathbf{r}(t) - \mathbf{r}(0)\right)^2 \right\rangle. \tag{1.29}$$

Note the resemblance between the variance of a population of diffusing particles and the mean square displacement of a single particle.

The MSD of a single diffusing particle satisfies the same relationship as the variance of a population of identical, non-interacting diffusing particles. In other

words, on long timescales, the MSD satisfies

$$\lim_{t\to\infty} \left\langle (\mathbf{r}(t) - \mathbf{r}(0))^2 \right\rangle = 6Dt. \tag{1.30}$$

This only holds on long timescales, as on short timescales particle motion is dominated by inertia. Inertial motion, also called ballistic motion, is not random, and thus doesn't lead to diffusive behaviour.

In summary, the average motion of a single particle undergoing Brownian motion in the absence of external fields is such that its MSD grows linearly in time. The rate at which this occurs is captured by the diffusion coefficient of the particle. In fact, this is true for any object undergoing Brownian motion. In particular, the center of mass (COM) of a polymer will also satisfy this relation in the absence of external fields. The relationship between the diffusion coefficient of the polymer's COM and that of its monomers will be explored in the next section.

1.3.2.3.2 Drift Velocity of a Free Monomer Now, consider the behaviour of a free monomer in a solvent when a constant, uniform external field $\mathbf{F}_{\text{ext}} = (F, 0, 0)$ is applied. Its equation of motion is now

$$m\mathbf{a} = -\gamma\mathbf{v} + \sqrt{2k_BT\gamma}\mathbf{R}(t) + \mathbf{F}_{\text{ext}}. \tag{1.31}$$

Consider the average behaviour of this monomer over long timescales. The time-average of the equation is

$$\lim_{t\to\infty} m \langle \mathbf{a} \rangle_t = \lim_{t\to\infty} \left(-\gamma \langle \mathbf{v} \rangle_t + \sqrt{2k_BT\gamma} \langle \mathbf{R}(t) \rangle_t + \mathbf{F}_{\text{ext}} \right) \tag{1.32}$$

$$= \lim_{t\to\infty} \left(-\gamma \langle \mathbf{v} \rangle_t + \mathbf{F}_{\text{ext}} \right), \tag{1.33}$$

since the random forcing term averages to zero. Since the average acceleration is the derivative of the average velocity, this is of the form

$$m\mathbf{u}' = -\gamma\mathbf{u} + \mathbf{F}_{\text{ext}}, \tag{1.34}$$

which is simply a terminal velocity problem for the average velocity. This terminal average velocity is solved for by setting the acceleration to zero. This yields

$$\lim_{t\to\infty} \langle \mathbf{v} \rangle_t = \frac{\mathbf{F}_{\text{ext}}}{\gamma}. \tag{1.35}$$

This is called the drift velocity of the monomer.

The drift velocity provides a more intuitive definition of γ: it is the ratio between the average terminal velocity of a particle and an applied external force field. This is also the basis for the definition of a particle's mobility μ. The mobility is given by

$$\mu = \frac{\lim_{t\to\infty} \langle \mathbf{v} \rangle_t}{\mathbf{F}_{\text{ext}}}, \tag{1.36}$$

the ratio of the drift velocity to the applied force. In other words,

$$\mu = \frac{1}{\gamma}. \tag{1.37}$$

These definitions of γ and μ suggest that they can be measured in experiments where an external force is applied to particles in solution, and the resulting drift velocities are measured. However, recalling that the diffusion coefficient satisfies

$$D = \frac{k_B T}{\gamma}, \tag{1.38}$$

this means

$$\gamma = \frac{k_B T}{D} \tag{1.39}$$

and

$$\mu = \frac{D}{k_B T}. \tag{1.40}$$

Using these relationships, γ and μ can also be found from experiments that measure the coefficient of diffusion at a known temperature. In particular, the mobility μ, which is defined as the ratio between drift velocity and applied force, can be measured in a diffusion experiment where no force is applied at all. This feature is called the Nernst-Einstein relation, and emphasizes that fluctuation effects (like the random solvent interactions leading to diffusion) are fundamentally coupled to dissipative effects (like the friction term leading to terminal velocities).

1.3.2.4 Coarse-Grained Polymers

The previous section shows how the second-order Langevin equation can be used to represent a single monomer in solvent. This section will create a polymer in the LD framework by defining multiple monomers and forces acting between them. In particular, this will be used to implement an approximation of a freely-jointed chain with excluded volume. The polymer thusly defined in the LD framework can then be used in numerical simulations.

Forces in physical systems are usually defined as the gradients of potential energies,

$$\mathbf{F} = -\overrightarrow{\nabla} U. \tag{1.41}$$

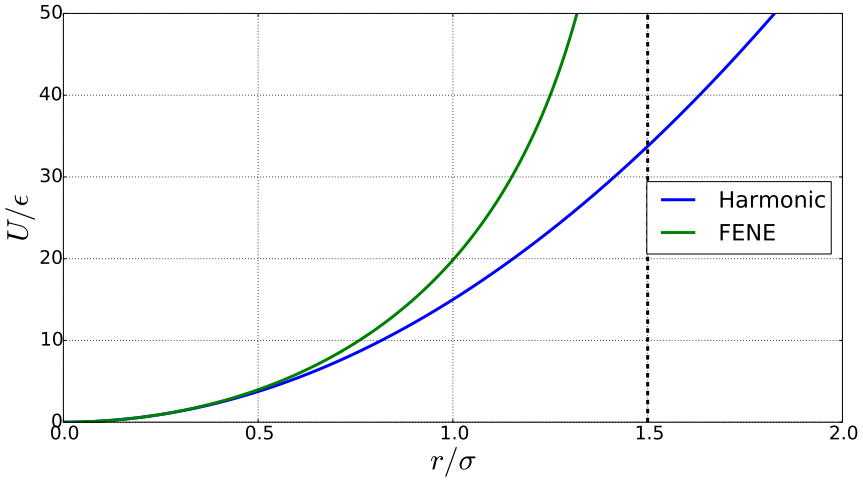

Figure 1.12: Comparison of the FENE potential with $\kappa = 30\epsilon/\sigma^2, r_{\max} = 1.5\sigma$ and an analogous harmonic potential $U_{\text{Harm}} = (1/2)\kappa r^2$.

For convenience, the forces introduced in this section will be discussed in terms of their corresponding potential energies. The units of energy will be expressed in terms of an energy scale parameter, ϵ. Similarly, distance will be expressed in units of σ, which corresponds roughly to the diameter of the hard sphere monomer model.

In the LD framework, the most natural way to bind monomers together is by defining attractive force terms that act between pairs of monomers. This thesis will use the finitely-extensible nonlinear elastic (FENE) potential to accomplish this. It is given by

$$U_{\text{FENE}}(r) = -\frac{1}{2}\kappa r_{\max}^2 \ln\left(1 - \frac{r^2}{r_{\max}^2}\right), \tag{1.42}$$

where r is the center-to-center distance between the two bonded monomers, κ is the FENE spring constant, and $r_{\max}$ is the maximum allowed bond length [41]. A plot of the potential is shown in Figure 1.12, using $\kappa = 30\epsilon/\sigma^2, r_{\max} = 1.5\sigma$.

At small bond lengths $r \ll r_{\max}$, the FENE potential resembles a harmonic

potential, since $x \ll 1 \implies \ln(1+x) \approx x$. Harmonic potentials are good approximations to many physical potentials, as they can be chosen to correspond to the second-order Taylor approximation of the correct potential. At larger bond lengths $r \approx r_{\max}$, however, the FENE potential grows more rapidly than a harmonic potential, and diverges to infinity at a finite bond length $r_{\max}$. This ensures that bond lengths can never exceed this finite length $r_{\max}$. The FENE potential is qualitatively a better model of polymer bonds than the harmonic potential: real polymers can only be stretched so much before undergoing chemical changes, such as severing the polymer. The FENE model captures a polymer that can be stretched a certain amount without such changes, and then requires an infinite amout of energy to stretch any farther. This is a consistent picture if the energies required to trigger chemical changes are much higher than the thermal energies of the system being considered. Figure 1.12 compares the FENE potential to an equivalent harmonic potential.

In the theoretical discussion of the freely-jointed chain with excluded volume, the monomers were treated as hard spheres. However, in the LD framework, a perfectly hard sphere corresponds to a discontinuous force: it is zero for $r \geq \sigma$ and infinite as soon as $r < \sigma$. Discontinuous forces create numerical errors or divergent solutions unless they are handled very carefully. It is far more convenient for numerical simulations to define the excluded volume as a continuous force. It must be zero for monomers that are far apart, then grow very rapidly as monomers become very close. This thesis will use the Weeks-Chandler-Andersen (WCA) potential for this purpose.

The WCA potential is a shifted and truncated version of the Lennard-Jones (LJ) potential. Both are shown in Figure 1.13. The LJ potential has the form

$$U_{\mathrm{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right], \tag{1.43}$$

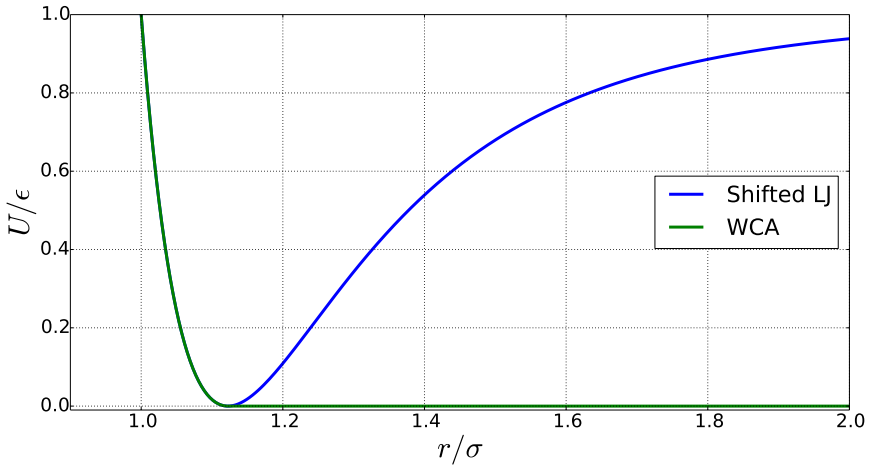

Figure 1.13: Comparison of the WCA potential and the LJ potential. The LJ potential is shifted upwards in energy by ϵ so that the shape of the two functions can be compared directly.

where r is the center-to-center distance between particles, ϵ is the energy scale, and σ is the distance scale. The LJ potential has a minimum value of $-\epsilon$ at $r = r_m = \sqrt[6]{2}\sigma$, diverges to positive infinity as $r \to 0$, and approaches 0 from below as $r \to \infty$. The term proportional to r^{-6} originates from the London dispersion force, which occurs between an instantaneous dipole (i.e. an atom with no net charge whose atomic orbitals have fluctuated in such a way as to form a non-zero dipole moment at some instant in time) and an induced dipole (i.e. the dipole moment resulting in a nearby neutral atom in response to the occurrence of an instantaneous dipole). As the electric field of each dipole goes as r^{-3}, the London dispersion force goes as r^{-6}. The term proportional to r^{-12} is used in this context to prevent the particles from overlapping. There is no precise physical origin for the functional form of the repulsive term. Rather, this term is simply meant to provide a strong repulsive force at short distances, and r^{-12} was chosen as it is rapid to compute once r^{-6} has been obtained (by squaring this value).

The WCA potential is derived from the LJ potential via

$$U_{\mathrm{WCA}}(r) = \begin{cases} U_{\mathrm{LJ}}(r) + \epsilon & r < r_c \\ 0 & r \geq r_c \end{cases}, \tag{1.44}$$

$$= \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon & r < r_c \\ 0 & r \geq r_c \end{cases}, \tag{1.45}$$

where r_c is called the cut-off distance. Setting $r_c = r_m$ ensures that both U_{WCA} and the associated force are continuous everywhere. The continuity of this force is important as discussed above. The continuity of the WCA potential itself is not necessary for the LD evolution of the system, but it is physically reasonable that the interaction energy between two particles should go continuously to zero as $r \to \infty$. Placing the cut-off at $r_c = r_m$ also means that the WCA potential leads to a repulsive force everywhere within the cut-off distance, followed by no force beyond the cut-off distance. This is our desired approximation to the hard-sphere monomer model.

1.3.2.4.1 Bond Crossing The FENE potentials bind monomers together, and the WCA potentials prevent monomers from overlapping. The combination is thus sufficient for an approximate implementation of a freely-jointed chain with excluded volume. However, one important modelling aspect remains: bond crossing. The values for the parameters of the FENE and WCA potentials will be chosen so as to resolve this issue.

Bond crossing, as the name suggests, occurs when the bond connecting two monomers A and B is allowed to pass through the bond connecting two different monomers C and D (Fig. 1.14). This violates the topology of the polymer, as the covalent bonds forming the backbone of the chain will never cross in reality. Bond crossing is clearly an important error when simulating systems that focus

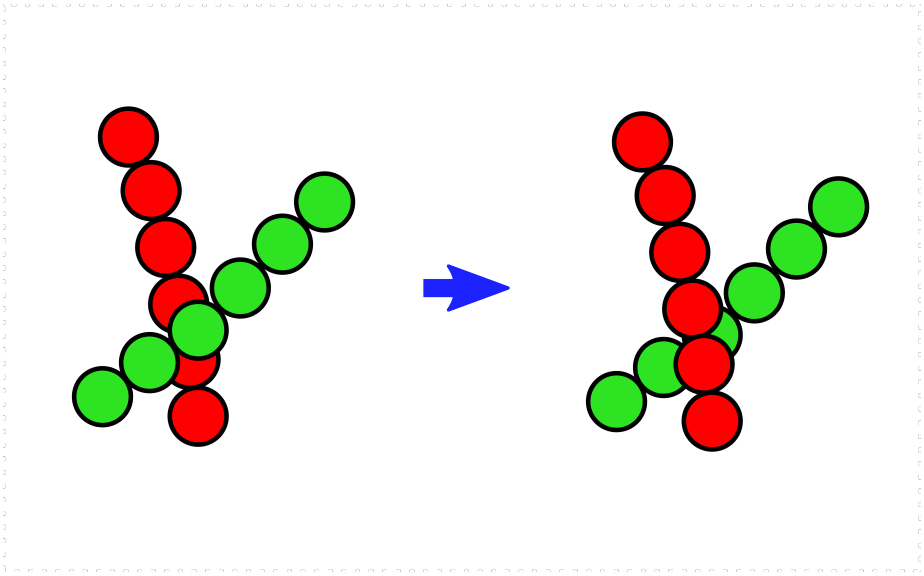

Figure 1.14: Illustration of bond crossing. The red and green monomers are part of the same polymer, but the red monomers occur far away from the green monomers along the backbone of the chain. If the polymer is not carefully implemented in simulation, the green monomers can slip between the bonds of the red monomers, violating the topology of the chain.

on topological features, like the evolution of knots in DNA. However, even in systems where topology is not obviously important, bond crossing greatly reduces the relaxation time of the polymer. It allows the chain to evolve easily between conformations that might otherwise take a very long time to access from one another. This may affect the outcome of systems where dynamical behaviour occurs on timescales much smaller than the (correct) relaxation time of the polymer.

By combining flexible bond lengths and excluded volume interactions correctly, bond crossing can be rendered sufficiently energetically unfavourable that it will never occur in simulations. The idea is to choose the FENE and WCA parameters so that adjacent monomers along the polymer are always close enough to one another that another pair of bonded monomers from elsewhere along the

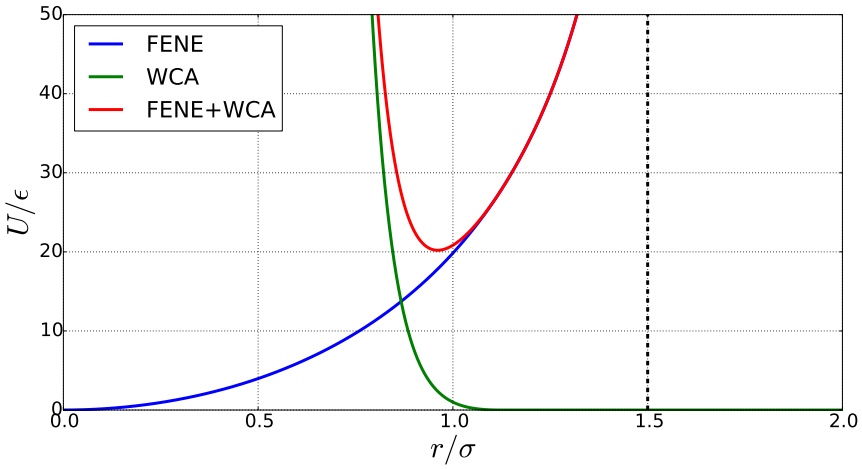

Figure 1.15: Plot of the total potential resulting from both FENE bonds and WCA excluded volume interactions between adjacent particles along a polymer.

polymer would require an exceedingly large amount of energy to cross bonds.

The parameters for the FENE and WCA potentials used for this study are chosen in accordance with the findings of Kremer and Grest [17]. The values of $\kappa = 30\epsilon/\sigma^2$ and $r_{\max} = 1.5\sigma$ lead to polymer bonds that are tight enough to prevent bond crossing, but are still generally numerically stable for common numerical integration schemes. In the current study, ϵ is set to $k_B T$, the thermal energy of the system. The combined potential is shown in Figure 1.15.

1.3.2.4.2 Hydrodynamic Interactions A coarse-grained LD implementation of a polymer is naturally an approximation to real polymer behaviour. The LD representation of the solvent does not capture information about the dynamics in the solvent itself. In particular, the drag and random force terms used to represent the solvent in LD act independently on each monomer. In reality, however, monomers that are in close proximity to one another will experience correlated interactions with the solvent. These correlations, due to the internal dynamics of the solvent, are referred to as hydrodynamic interactions

(HI) [73].

As an illustration of the effects of HI, consider the dynamics of our polymer model in the absence of external forces. As alluded to in the previous section, the COM of the polymer will diffuse in the solvent with some effective diffusion coefficient. There are two analytic theories commonly used to relate the center-of-mass diffusion of a polymer in free solution to the diffusion of its constituent monomers.

The first model, called the Rouse model, neglects hydrodynamic interactions between the various portions of the chain. It predicts

$$\gamma_{\text{polymer}}^{\text{Rouse}} = N\gamma_{\text{monomer}}, \tag{1.46}$$

where γ is the friction coefficient [73]. In other words, the total drag force on a polymer made of N monomers is simply the sum of the drags on each monomer, since the drags all act independently in the Rouse model.

The second model, called the Zimm model, accounts for these neglected hydrodynamic interactions, and predicts

$$\gamma_{\text{polymer}}^{\text{Zimm}} = N^{\nu}\gamma_{\text{monomer}}. \tag{1.47}$$

where $\nu \approx 0.588$ is the Fleury coefficient [73]. Since $\nu < 1$, the drag on a Zimm polymer is less than the drag on a Rouse polymer at large values of N. This occurs because, as the polymer moves, the solvent near the polymer begins to move along with it. The relative velocity between the solvent and the polymer is reduced by this, which in turn reduces the drag.

In both cases, $D = k_B T/\gamma$. In other words, long polymers diffuse much more rapidly under the Zimm model than the Rouse model.

In reality, the Zimm model is a more accurate description of polymer diffu-

sion in free solution in the absence of external forces. However, in the presence of an electric field, the flow of ions through the solution disrupts hydrodynamic correlations between the various portions of the chain. In this regime, Rouse dynamics describe polymer diffusion more accurately. Indeed, this is the explanation for the so-called free-draining solution result of polymer electrophoresis: the drift velocity of Rouse polymers experiencing a uniform force field acting with a magnitude F on each monomer is $v = (NF)/(N\gamma) = F/\gamma$, which is independent of chain length. Thus a solution of uniformly-charged polymers in free solution cannot be separated by length by applying an electric field.

In a nanopore system, polymers far from the nanopore experience very little electric field, and thus might be expected to diffuse according to the Zimm model. Conversely, near the nanopore, the electric field is strong enough that the Rouse model might be more appropriate. Furthermore, this implies that the polymer diffusion in some region in between "far from the nanopore" and "close to the nanopore" will be described by some combination of the two models. The correct choice of diffusion parameters for the microscopic model is therefore unclear, and will be the subject of further study.

These predictions for the diffusion coefficients of the COMs of polymers can be used to derive theoretical estimates for the relaxation time of polymers. In the previous section, the relaxation time was introduced as the timescale over which the autocorrelation of the polymer's end-to-end length decays exponentially. Now, theoretical estimates will be obtained as the timescale over which a polymer is expected to diffuse over a distance equal to its own size. In particular, the condition is

$$\mathrm{MSD}(\tau_R) \approx R_G^2. \tag{1.48}$$

Using the Rouse model, this is

$$6\frac{D}{N}\tau_R \propto N^{2\nu} \implies \tau_R^{\text{Rouse}} \propto N^{2\nu+1}. \tag{1.49}$$

Using the Zimm model, this is

$$6\frac{D}{N^{\nu}}\tau_R \propto N^{2\nu} \implies \tau_R^{\text{Zimm}} \propto N^{3\nu}. \tag{1.50}$$

1.3.2.4.3 Electrostatic Interactions As discussed previously, the polymer translocation studied in this thesis is driven by the application of an electric field across the nanopore. This presumes that the polymer carries a charge. Indeed, in the generic polymer model introduced here, each monomer will be assumed to carry an identical unit charge. Conversely, the only electrostatic interaction that is modelled explicitly is the force exerted by the electric field on the monomers.

The electrostatic interaction between monomers is shielded by the ions in the electrolytic solvent. Any region of the polymer that carries a net charge causes the ions in the solvent to rearrange so as to minimize the local electrostatic energy. In particular, ions of opposite charge are attracted, and those of like charge are repelled. This leads to the formation of what is called an ion cloud around the charges of the polymer.

These clouds reduce the net charge of the monomers in such a way that the net electrostatic force between monomers decays exponentially with the distance between them. The length scale of this decay is known as the Debye length. This study assumes that the Debye length is short enough that these monomer-monomer electrostatic interactions can be treated as short-range effects. Indeed, these interactions are subsumed into the excluded volume interaction, and otherwise neglected.

This shielding effect is an example of electrohydrodynamics (EHD). Another EHD effect is the disruption of hydrodynamics interactions by the flow of ions. This causes polymer to behave like Rouse polymers in regions of strong electric fields, as discussed above. Indeed, this phenomenon means the HIs discussed above are generally reduced inside the nanopore.

Other, more complicated EHD effects are neglected for the purpose of this preliminary characterization of the cavity-nanopore device. Although such interactions can certainly affect translocation dynamics, they are not expected to compromise the fundamental results of this study. Future work will explore the implications of more complicated EHD effects on CN translocation.

1.3.2.5 Numerical Integration

The previous sections have established a polymer model in the LD framework. The dynamics of the polymer are determined by solving the coupled second-order Langevin equations for all the monomers simultaneously. The present model for a chain of N monomers consists of N coupled second-order stochastic differential equations, which are non-linear in general. In other words, it is far too complicated to be solved directly. Instead, the LD equations of motion are numerically integrated.

Numerical integration is used in this case to obtain the position of a particle from the LD equations of motion, which specify only its acceleration. The integral of acceleration over time is the change in the particle's velocity; the integral of velocity is the change in position. However, the equations of motion are too complicated to integrate analytically. By using Taylor series approximations of the particle's position, velocity, and acceleration, one can derive algorithms to estimate the integral of the equation of motion numerically over short intervals of time. At each moment in time, the forces acting on each of the monomers are calculated, and this numerical integration algorithm is used to update all of

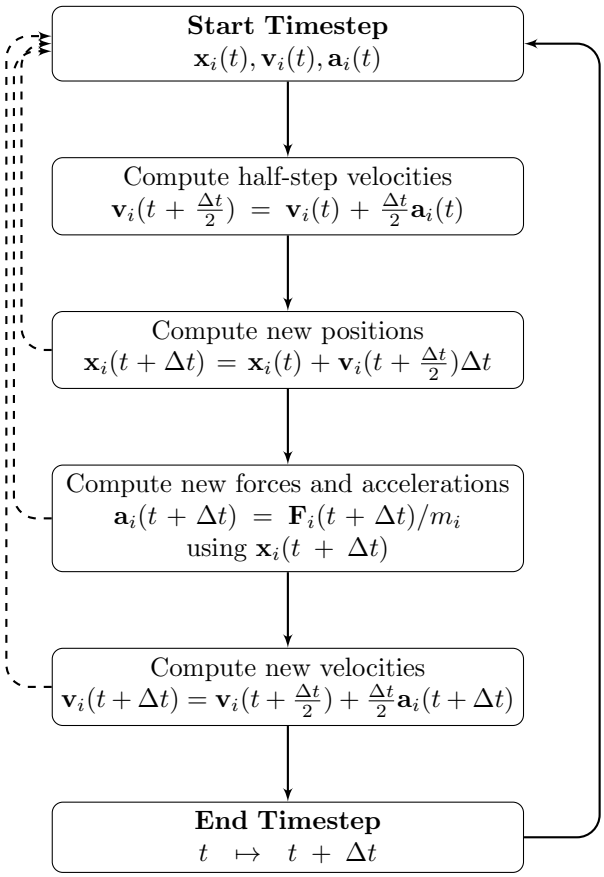

Figure 1.16: Flow chart of the Velocity Verlet algorithm.

their positions.

1.3.2.5.1 Velocity Verlet The numerical integration algorithm used in this thesis is the Velocity Verlet algorithm [75]. It is described below, and illustrated by a flow chart in Figure 1.16. It keeps track of the position and velocity of each monomer at each point in time. It integrates forwards in time by intervals of Δt, called the timestep size.

1. At time t, calculate the net force acting on each monomer i, $\mathbf{F}_i(t)$, using the current $\mathbf{x}_i(t)$ and $\mathbf{v}_i(t)$. Use this to calculate the acceleration of each

monomer, $\mathbf{a}_i(t) = \mathbf{F}_i(t)/m_i$.

2. Calculate the so-called half-step velocities, $\mathbf{v}_i(t+\frac{1}{2}\Delta t) = \mathbf{v}_i(t)+\frac{1}{2}\mathbf{a}_i(t)\Delta t$ for each monomer.

3. Update the positions to $\mathbf{x}_i(t+\Delta t) = \mathbf{x}_i(t) + \mathbf{v}_i(t+\frac{1}{2}\Delta t)\Delta t$ for each monomer.

4. Using the new monomer positions, re-calculate the net forces and accelerations of all the monomers at $t+\Delta t$.

5. Update the velocities of the monomers to $\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t+\frac{1}{2}\Delta t) + \frac{1}{2}\mathbf{a}_i(t+\Delta t)\Delta t$.

6. Repeat.

The algorithm above requires one to compute the net force on each particle at each timestep, but the random forcing term is defined as a continuous random process in time. In order to account for this random force correctly at the discrete timesteps of this algorithm, it must be computed as

$$\mathbf{F}_{\text{random}}(t) \approx \sqrt{\frac{2k_BT\gamma}{\Delta t}}\mathbf{R}(t). \tag{1.51}$$

The derivation of this form is beyond the scope of this thesis.

In this thesis, the Velocity Verlet algorithm was used through the open-source ESPResSo molecular dynamics software package [41]. Using ESPResSo has many advantages compared to implementing a numerical solver manually. Given its larger user base, ESPResSo has quality assurance testing that cannot be matched by personal solver implementation. Also, ESPResSo comes with a variety of optimizations and useful functions, including neighbour list algorithms and cylindrical pore geometric constraint objects. Finally, the members of the

cNAB.LAB have years of experience using ESPResSo, providing a invaluable knowledge base.

1.3.2.5.2 Neighbour Lists In neighbour list algorithms, the numerical integration scheme keeps track of which LD particles are close to which others: this list is stored as a list of neighbours for each LD particle [75]. The numerical integrator will then only compute interactions between neighbours. The neighbour lists are updated periodically to account for particle motion. The use of neighbour lists reduces the number of interactions in a system of N particles from $\mathcal{O}(N^2)$, where every particle interacts with every other particle, to $\mathcal{O}(NN_N)$, where N_N is the average number of neighbours around each LD particle. In diffuse systems, N_N will remain roughly constant with N (being a function of local density only), so that the total number of interactions that must be computed at each integration step is greatly reduced in large systems. The actual computational cost of a numerical integration scheme using neighbour lists must also account for the cost of updating the neighbour lists every so many integration steps, but the net effect is still a significant improvement in performance for many LD systems.

1.3.2.6 Boundary Conditions

In the nanopore systems being studied here, the polymer must interact with a thin membrane (made of e.g. SiN). In reality, this membrane will be made of an array of atoms. However, the atomic detail is unnecessary given the scale at which the polymer is coarse-grained. For this polymer model, the membrane will be approximated as a uniform plane with a small cylindrical hole in it.

The simplest model of interaction between monomers and this plane would be to treat the plane as a hard reflecting wall. This is analogous to the hard sphere approximation used in the theoretical discussion of the excluded volume.

However, as discussed above, using this hard wall model results in a force that is a discontinuous function of monomer position. As with the hard sphere model, it is more convenient to approximate the hard walls by the WCA potential.

In the ESPResSo software package, planar surfaces with a cylindrical hole in them can be defined using the pore constraint objects [41]. They are defined by the following parameters:

- The coordinate of the center of the pore.
- The normal vector of the plane.
- The nominal radius and nominal half-length of the pore.
- The effective particle type of the membrane.

In order to incorporate this boundary into the LD equations, ESPResSo treats the membrane as if it were constructed out of an infinite number of particles. When a real LD particle approaches the membrane, ESPResSo determines the point P on the surface that is closest to the particle. It then generates a force on the particle as if an imaginary particle were centered at P. The nature of the force produced by the imaginary particle is specified by defining the effective particle type of this imaginary particle. In this thesis, the same WCA potential that acts between the monomers that form the polymer, Equation 1.45, acts between the monomers and the imaginary particles forming the membrane.

Figure 1.17 illustrates the geometry of the pore constraint as used in ESPResSo, and how these constraints are used here to define the geometry of the cavity-nanopore system. The values r_{eff} and t_{eff} are known as the effective radius and effective thickness, respectively. These effective quantities account for the excluded volume interaction between the polymer and the wall. If the WCA potential is treated approximately as a solid boundary a distance $\sigma/2$ from the

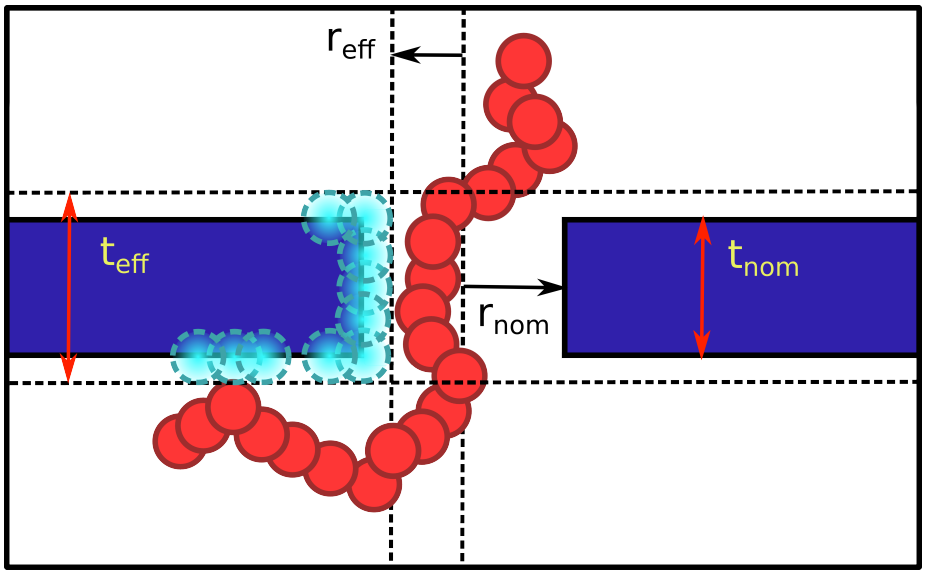

Figure 1.17: Illustration of the pore constraint used in ESPResSo. The definition given by the user in the ESPResSo user manual must specify the nominal radius, r_{nom}, and the nominal half-thickness, $l_{\text{nom}} = t_{\text{nom}}/2$.

center of each particle, then the walls can be approximated as solid walls located at the effective radius and thickness.

1.4 Literature Review

1.4.1 Experimental Studies of Standard Nanopores

The earliest proposal to use nanopores as DNA sequencing tools was in 1996 by Kasianowicz et al. [33]. A recent review paper by Feng et al. covers experimental advances in nanopore technology and summarizes the state-of-the-art in 2016 [13]. The review fails to address the work by Kwok et al. that may reduce the manufacturing costs of solid-state nanopores dramatically [37]. It uses a controlled dielectric breakdown of a solid-state membrane to create a nanopore, whereas older methods use expensive electron or ion accelerators to drill holes

into such membranes [37, 13].

The nanopore concept bears close resemblance to Coulter counters, which date back decades before modern nanopore technologies [20, 13]. In a Coulter counter, small analytes in solution are forced through a hole of comparable size using an electric field, and the ionic current flowing through the hole is measured [20]. Clearly this is closely analogous to the measurement techniques in nanopore translocation experiments, although polymer analytes are significantly more complicated than the simpler analytes for which Coulter counters were originally conceived.

The main challenge remaining in nanopore-based DNA sequencing is a matter of spatial and temporal resolution of the base pairs [13]. Currently, the only commercially available nanopore-based sequencing device, the MinION (described previously), still has an error rate of roughly 38.2% [39]. Slowing down translocation will enable more precise characterization of DNA molecules during translocation [14].

1.4.2 Theoretical and Simulation-Based Studies of Standard Nanopores

Explaining the experimental results of the previous section with theoretical models or in numerical simulations has been the subject of hundreds of studies. Many literature reviews have been published, of which those by Panja et al. [54] and Palyulin et al. [49] are most focused on material relevant to simulation work. This section will start by reviewing results for unbiased translocation, which is polymer translocation in the absence of an external driving force, before reviewing results for translocation driven by an external electric field. The section will conclude with a review of simulation studies exploring exotic nanopore geometries, especially those that resemble the cavity-nanopore geometry that is

studied in this thesis.

1.4.2.1 Unbiased Translocation

Unbiased translocation is translocation that occurs spontaneously, without any external driving force. The following derivation, following that by Panja et al. [54], demonstrates that unbiased translocation occurs at a negligble rate in practice. The polymer must overcome a considerable free energy barrier for this to occur. In free solution, the multiplicity of polymer states for a chain of N monomers scales as

$$Z_b(N) \sim A\mu^N N^{\gamma-1}, \tag{1.52}$$

where $\gamma \approx 1.16$ for all three-dimensional polymers, while A and μ depend on the details of the polymer [54].

Instead of a free polymer, consider now a polymer with one end fixed against a membrane. This scales as

$$Z_w(N) \sim A_1\mu^N N^{\gamma_1-1}, \tag{1.53}$$

where μ is the same as in Equation 1.52, but γ_1 and A_1 differ from γ and A [54]. In particular $\gamma_1 \approx 0.68$ for three-dimensional polymers, so $\gamma_1 \ll \gamma$.

If a polymer spontaneously translocates more than half-way, then it is more likely to cross the membrane to the other side than it is to return to its original side. Thus the free energy barrier to spontaneous translocation can be estimated using the difference between $Z_b(N)$, the multiplicity of a free polymer, and the multiplicity of a polymer that has translocated half its monomers. Assuming the membrane is thin enough that no monomers are inside the pore, a half-way translocated chain has $N/2$ monomers on each side of the membrane.

So its multiplicity is $Z_w(N/2)$ per side, or $Z_w^2(N/2)$ in total. Since unbiased translocation has no external fields, the free energy of the polymer is simply its entropy, and the difference in entropy between free and half-way translocated states is

$$\Delta S \sim \ln(Z_b(N)) - \ln(Z_w^2(N/2)) \tag{1.54}$$

$$= \ln \frac{Z_b(N)}{Z_w^2(N/2)} \tag{1.55}$$

$$= (\gamma - 2\gamma_1 + 1)\ln(N) + \ln \frac{A4^{\gamma_1 - 1}}{A_1^2} \tag{1.56}$$

$$\sim 0.8 \ln(N), \tag{1.57}$$

where the coefficient of $\ln(N)$ is roughly 0.8 from the above values of γ, γ_1 for three-dimensional polymers. Thus sufficiently long polymers must overcome a large free energy barrier for spontaneous translocation to occur.

Although unbiased translocation occurs at negligible rates in reality, its study yields important insight into nanopore translocation dynamics. The earliest theoretical work on unbiased translocation was conducted by Sung and Park in 1996 [72]. They considered only the quasi-equilibrium translocation process (where the polymer is relaxed at all points during the translocation process) of a very long chain, $N \gg 1$, and assumed the capture process could be ignored for this analyis. The free energy of the translocating chain can be expressed in terms of the number of monomers remaining on the *cis* side, m. Using the form of $Z_w(N)$ given above, the multiplicity of the polymer when m monomers remain on *cis* is $Z_w(m)Z_w(N-m)$, so its free energy is proportional to $\ln[m(N-m)]$.

In the long chain limit, this means the free energy of the polymer is very flat around $m = N/2$. The force resulting from the free energy gradient is thus very small, and so quasi-static translocation is a diffusive process [54]. Recall Equation 1.30 for the MSD of a diffusive process. The timescale for transloca-

tion to finish is that required for m to diffuse the length of the chain, which is proportional to N. Thus the quasi-static picture of unbiased translocation predicts that the translocation time will scale as

$$\tau \sim N^2/D, \tag{1.58}$$

where D is the effective diffusion coefficient of m. Sung and Park used $D \sim N^{-1}$ for Rouse polymers and $D \sim N^{-1/2}$ Zimm polymers, and obtained $\tau \sim N^{\alpha}$ with $\alpha = 3$ and $\alpha = 5/2$, respectively [72].

Whereas Sung and Park had considered a phantom polymer (i.e. one without excluded volume interactions), Muthukumar conducted the same analysis for self-avoiding chains (i.e. chains with excluded volume interations) [48]. Furthermore, he argued that the diffusion coefficient of m should not depend on N, and found $\alpha = 2$ [48]. However, both pictures are incorrect, as unbiased translocation is in fact not an equilibrium process for sufficiently long chains, as demonstrated below.

The first to propose that polymer translocation cannot be modelled as a quasi-static proces was Chuang et al. [6]. The argument is simple: if translocation is quasi-static, then the chain must be in a relaxed conformation at each point in translocation. The relaxation time of, for instance, a Rouse polymer with excluded volume scales as $\tau_R \sim N^{1+2\nu}$ (Equation 1.49). However, the quasi-static analysis of Muthumukar predicted that the translocation time scales as $\tau \sim N^2$ for such polymers [48]. For large N, then, the relaxation time will be longer than the translocation time, which contradicts the assumption that the chain can be relaxed at each point during translocation. Thus the quasi-static picture of translocation is inconsistent with itself, and cannot be correct.

Using this line of reasoning, Chuang et al. also proposed a theoretical lower bound for the fastest possible translocation time [6]. The fastest unbiased

translocation will occur for nanopores so wide that the polymer can pass through the pore completely unhindered. In this case, translocation is simply free diffusion: the polymer must simply diffuse across the hole in the membrane. The distance it must travel is roughly its own radius of gyration, R_G, if it starts right beside the membrane on the *cis* side and finishes right beside the membrane on the *trans* side (and the membrane is thin). This is the same condition used to derive Equations 1.49 and 1.50 for the relaxation time of polymers, and indeed the scaling relation is the same. In particular, the fastest possible scaling for the translocation of a Rouse polymer is, by this model,

$$\tau \gtrsim N^{1+2\nu} \quad [6]. \tag{1.59}$$

Furthermore, simulations conducted in the same paper found that this lower bound was saturated, i.e. that the translocation time indeed scales as $\tau \sim N^{1+2\nu}$ [6].

Following the publication of this approach to setting a lower bound on the scaling of translocation time, an enormous number of studies were conducted to test when this formalism applied. This body of literature is colloquially referred to as "the exponent wars", as they centered around various approaches to measuring α in the relationship $\tau \sim N^{\alpha}$. Several simulation studies published agreement with the approach of Chuang et al. [6], using various simulation techniques [45, 26, 76].

Conversely, later simulation studies found results inconsistent with Chuang et al, indicating rather $\alpha = 2 + \nu$ for Rouse polymers and $\alpha = 1 + 2\nu$ for Zimm polymers [52, 51, 53, 10, 15, 31, 8, 12]. A theoretical explanation of these results was published by Panja et al. [52, 51, 53, 50]. The theory postulates that when a monomer crosses the membrane, it creates a local strain in the polymer bonds. The strain can either be relaxed by propagating along the

polymer's backbone, or if the monomer hops back across the membrane. The propagation of strain along the polymer's backbone occurs on timescales of τ_R, the polymer's relaxation time. However, including the option to resolve local strain propagation by crossing the membrane leads to predictions of $\alpha = 2 + \nu$ for Rouse polymers and $\alpha = 1 + 2\nu$ for Zimm polymers [52, 51, 53, 50]. This model will be referred to as the memory function approach.

The exponent wars were eventually resolved by de Haan and Slater [9]. Using a simulation methodology very similar to that described in this thesis, they studied the effects of varying the viscosity of the solvent, η (which is captured in LD via the parameter γ) [9]. The results of that study [9] are greatly consistent with the theoretical memory function approach [52, 51, 53, 50]. At $\eta = 0$, the bond strain relaxes instantaneously, and both the memory function theory and the simulation results of de Haan and Slater obtain $\alpha = 2$ [52, 51, 53, 50, 9]. As η is increased, α increases, crossing $\alpha = 1 + 2\nu$, and appearing to converge to $\alpha = 2 + \nu$ at the highest viscosities accessible to simulations [52, 51, 53, 50, 9].

Thus, the work of de Haan and Slater [9] demonstrates that the relation $\tau \sim N^\alpha$ does not hold for a single value of α across all translocation conditions. In particular, α can appear to hold different values depending on the simulation conditions. However, their work also showed that if τ/N^2 is plotted against ηN^x with $x = 0.516$, then all of the published scaling results for simulations of unbiased translocation collapse to a single line. This universal curve suggests that $\alpha \to 2.516$ at high viscosities, which is in agreement with $\alpha = 2 + \nu$ predicted by the memory function approach [52, 51, 53, 50, 9].

1.4.2.2 Driven Translocation

Driven translocation, as introduced earlier, involves adding an external force to the unbiased translocation experiment. This external force provides the activation energy required to initiate translocation, since there is an entropic barrier

to spontaneous translocation. In addition to this, the external force will change the scaling of the translocation time τ with the chain length N. This discussion will assume that the external force is constant for the duration of translocation.

The simplest approach to predicting the scaling of τ with N in driven translocation starts with the quasi-static approximation, as was the case for unbiased translocation. In unbiased translocation, the free energy function was very flat near $m = N/2$, where m is the number of monomers remaining on the *cis* side. From this, it was argued that the entropic force driving translocation was very weak for most of the translocation, so that it was a predominantly diffusive process. Conversely, in biased translocation, the external force field tilts the entire free energy landscape to one side. For large external forces, then, if biased translocation is a quasi-static process, then it is dominated by drift, rather than diffusion. This approach therefore predicts $\tau \sim N^{-1}$, the time it takes for m to drift at a constant velocity across the entire length of the chain [48, 44]. Similarly, since the drift velocity in an overdamped system is linear in the applied force F, this implies $\tau \sim F^{-1}$ [48, 44].

Naturally, since unbiased translocation was shown to occur too rapidly to be quasi-static, it seems implausible that the (faster) driven translocation would be quasi-static. This was first shown by Kantor and Kardar [30], again by considering the limiting case of an infinitely wide pore, in this case for a Rouse polymer with excluded volume. For a constant force F, they argued that the center of mass of the polymer attains a terminal velocity proportional to $v_{\text{drift}} \sim F/N$, in accordance with Equation 1.46, which states that the coefficient of friction scales with N for Rouse polymers. As in the case of unbiased translocation through an infinitely wide pore, translocation is complete when the polymer's center of mass travels a distance on the order of its radius of gyration, $R_G \sim N^{\nu}$. This means that $\tau \sim R_G/v_{\text{drift}} \sim N^{1+\nu}/F$ [30].

This again implies that the translocation time in the wide pore limit, $\tau \sim N^{1+\nu}$, is faster than the relaxation time of the same polymer, $\tau_R \sim N^{1+2\nu}$ (Equation 1.49), so Kantor and Kardar proposed this as a lower bound for translocation time, like Chuang et al. did for unbiased translocation [6, 30]. Similarly, Kantor and Kardar also conducted simulations to test this bound, and found that it was saturated. This again led to a huge body of literature testing this lower bound: another so-called exponent war [54]. This second exponent war was even more dynamic than the first, as the addition of a driving field adds another layer of complexity to the system.

The resolution of this second exponent war can be attributed to a theoretical picture that captures the dynamics of the polymer during translocation. The idea, first proposed by Sakaue et al. [61], is that if the polymer is at equilibrium at the beginning of translocation, the driving force only perturbs the equilibrium of the portions of the chain near the nanopore. The perturbation propagates down the polymer chain as a tension wave in the monomer-monomer bonds. It travels with a finite speed, so that only a finite portion of the chain is within the tension front during the first part of translocation.

This so-called tension propagation model was refined by Sakaue et al. [62, 59, 60, 63], and was eventually used by Ikonen et al. [28, 27] to unify simulation results from several disparate methodologies. The final model considers three translocation regimes: very strong driving force, very weak driving force, and the intermediate force regime. Ikonen et al. found that pore-polymer interactions and finite chain length effects have very important effects on translocation dynamics [28]. In particular, many theoretical models considered the long-chain limit, whereas Ikonen et al. demonstrated that both experiments and simulations consider chains that are too small to be approximated by this limit [28]. When all of these effects are included, they recover excellent agreement with

published results. In particular, the long-chain limit scaling is $\tau \sim N^{1+\nu}$, but under realistic experimental and simulation conditions this is not achieved, and translocation time does not obey a relation of the form $\tau \sim N^{\alpha}$ (or, worded differently, α depends strongly on the details of the experimental procedure or simulation parameters) [28].

1.4.2.3 Exotic Nanopore Geometries

The previous sections have reviewed the state of the literature for the standard nanopore geometry, namely a cylindrical hole of constant radius through a thin membrane. This thesis, on the other hand, characterizes polymer translocation through a novel nanopore geometry, where the inside of the nanopore contains a cavity of larger radius. This section reviews studies of similar, exotic nanopore geometries (i.e. geometries that differ from the standard one). In particular, the following nanopore geometries all involve some form of confinement of the polymer between the beginning and end of translocation.

Langecker et al. conducted an experimental investigation of a system composed of two consecutive nanopores stacked on one another [38]. The first nanopore was of the standard geometry, whereas the second nanopore had a tapered conical geometry. This type of stacked conical nanopore geometry is illustrated in Figure 1.18. The geometry is somewhat analogous to the cavity-nanopore: translocating polymers are rather confined at the entrance and exit of the device, but relatively unconfined in between. However, the device constructed by Langecker et al. was much larger than the radius of gyration of the DNA strands with which they experimented: their device was 23 nm at its narrowest point, and 1.5 μm at its widest point [38]. Conversely, they used 10 kbp DNA strands with a contour length of 3.4 μm [38]. The radius of gyration of the DNA strands depends on the electrolytic conditions, which they varied, but will always be significantly less than the contour length, so that in this device

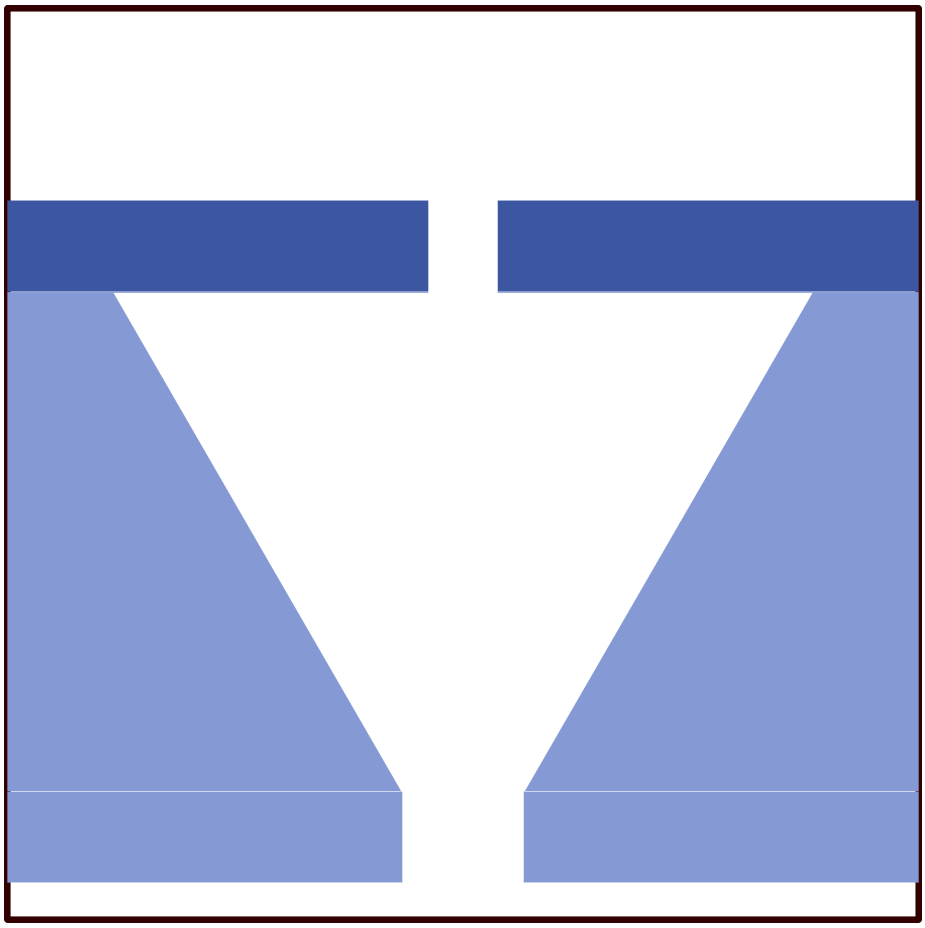

Figure 1.18: Illustration of the stacked conical nanopore geometry.

the radius of the confinement was much larger than the radius of gyration [38].

Pedone et al. studied the same geometry used by Langecker et al. in terms of the diffusion of spherical nanoparticles [55]. Similarly, Harms et al. studied the transport of viral capsids between two standard nanopores in series held 2 μm apart [21]. None of these three studies (i.e. Langecker et al. [38], Pedone et al. [55], and Harms et al. [21]) witnessed the dynamics described in this thesis for the cavity-nanopore system, because these dynamics are unique to polymer translocating through system where the radius of confinement is comparable to

the radius of gyration of the polymer.

Conversely, translocation from α-hemolysin (Fig. 1.1) has also been studied. For instance, Sun et al. simulated polymer translocation through α-hemolysin using a similar methodology to that presented in this thesis [71]. The structure of α-hemolysin involves a highly confining barrel structure, much narrower than the radius of gyration of the translocation. In other words, whereas the experimental studies listed above had a confinement much larger than R_G, the confinement in α-hemolysin is much smaller than R_G. As such, the study by Sun et al. also did not capture the dynamics presented here for the cavity-nanopore system [71].

A recent experimental study by Liu et al. did indeed have a confining geometry where the radius of confinement was comparable to the radius of gyration of the polymers they used [43]. However, the exit nanopore from their geometry was much larger than the entrance nanopore [43]. Furthermore, and more importantly, they did not conduct complete translocation experiments: an electric field was applied only long enough to cause the polymer to enter the cavity, and then the field was turned off, so that the polymer remained trapped inside the cavity [43]. Thus, this set of experiments did not capture the translocation dynamics found in the current study of the cavity-nanopore. However, this work demonstrates the experimental feasibility and relevance of the cavity-nanopore device.

Finally, a simulation study by Mökkönen et al. used an LD simulation methodology similar to the one used in this thesis to explore polymer escape from a two-dimensional confinement [47]. Although this is not a translocation study, the study reported the escape rate from the confinement as a function of chain length, which was found to be a non-monotonic function of chain length [47]. However, whereas this thesis report a minimal translocation time for chains

of intermediate length, the study by Mökkönen et al. reports a minimal escape rate for chains of intermediate length, which is indeed the opposite result [47]. This can be explained by understanding that the regime of chain lengths studied by Mökkönen et al. is much smaller compared to the dimensions of their confining potential than the chains in the present thesis [47]. In their long-chain limit, for instance in Figure 3(b) of their paper, a maximum in the escape rate is apparent [47]. This maximum escape rate corresponds to the minimal translocation time found here.

Chapter 2

Results

2.1 Simulation Setup

This section describes a new nanopore design, the cavity-nanopore (CN). The system is modelled in the CGLD formalism described in the previous chapter. Simulations are implemented using the ESPResSo software package [41].

2.1.1 The Cavity-Nanopore

Figure 2.1 shows an illustration of the CN design under study here. The CN design can be described as a long SN in which a large cavity has been introduced between the entrance and exit. The hole through which polymers enter the CN will be referred to as the *cis* or entrance pore, and the hole through which they exit will be referred to as the *trans* or exit pore. For simplicity, the radius of these two pores will be made equal. The space between the entrance and exit pores will be referred to as the cavity.

The purpose of this thesis is to characterize the dynamics of polymer translocation through the CN. In order to draw comparisons to the extensive SN liter-

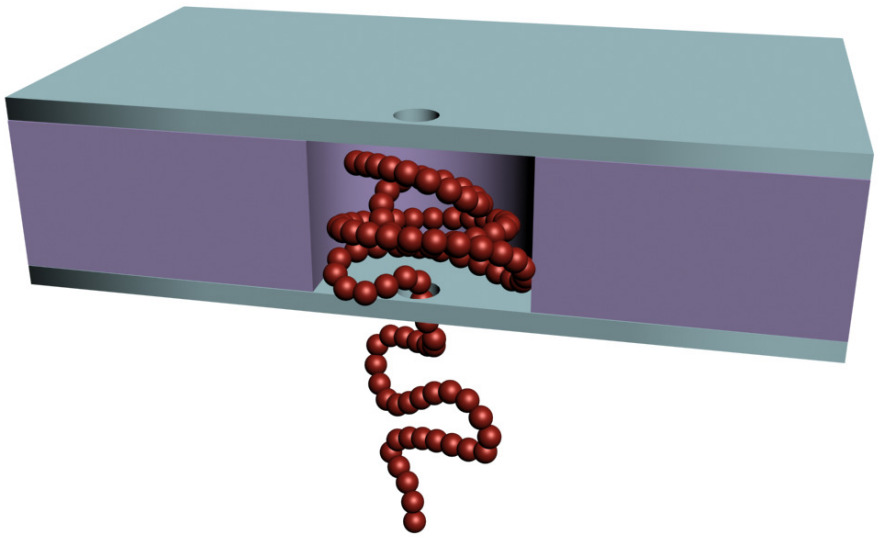

Figure 2.1: Illustration of a polymer entering the cavity-nanopore design.

ature, SN translocation simulations were also conducted. The SN geometry for these simulations was chosen equal to the geometry of the entrance pore of the CN, i.e. the CN with the cavity and exit pore membranes deleted.

The simulations were conducted using the CGLD polymer model described in the Introduction. The walls of the nanopores were modelled using the pore constraint boundary conditions. Simulations were conducted using the ESPResSo software package, which contains implementations of these models and uses the Velocity Verlet algorithm with neighbour lists to calculate polymer dynamics [41].

Translocation was driven by an electric field, which is described in the next section. The initial conditions of the simulation are described after that. Finally, the simulation parameters for which data were collected are summarized before the simulation results are presented.

2.1.2 Software and Computer Systems

The simulation work presented in this thesis was conducted on SHARCNET, the Shared Hierarchical Academic Research Computing Network [1]. Simulations were run using the ESPResSo molecular dynamics software package [41]. Plotting was conducted using the matplotlib package in Python [25]. Visualization of particle trajectories was conducted using VMD [24].

2.1.3 Electric Field

Polymer translocation in the simulations was driven by an electric field acting across the membrane. The field in the simulations is designed to model that which would be generated in nanopore experiments, where a fixed voltage is applied across the system by electrodes held far from the nanopore. The membrane in which the nanopore exists is a good insulator, and the nanopore is much smaller than the membrane. This arrangement results in an electric field that is very strong near the pore and very weak far from the pore.

This study will use a minimal model of the electric field. The electric field will be approximated as being parallel to the axis of the nanopore at all points in the system. Furthermore, it will be taken to depend only on the axial coordinate. If the axis of the nanopore is chosen to be the z axis, this means the field is of the form $\mathbf{E}(x, y, z) = (0, 0, E(z))$ everywhere. The relative magnitude of the field at each point in the system will be determined by the conservation of electric flux. The resulting field, shown in Figure 2.2, is zero outside the CN, strong in the entrance and exit pores, and weak inside the cavity.

Consider the electric flux passing through an axial segment of the system located at z, i.e. an infinite plane normal to the axis of the nanopore. This flux

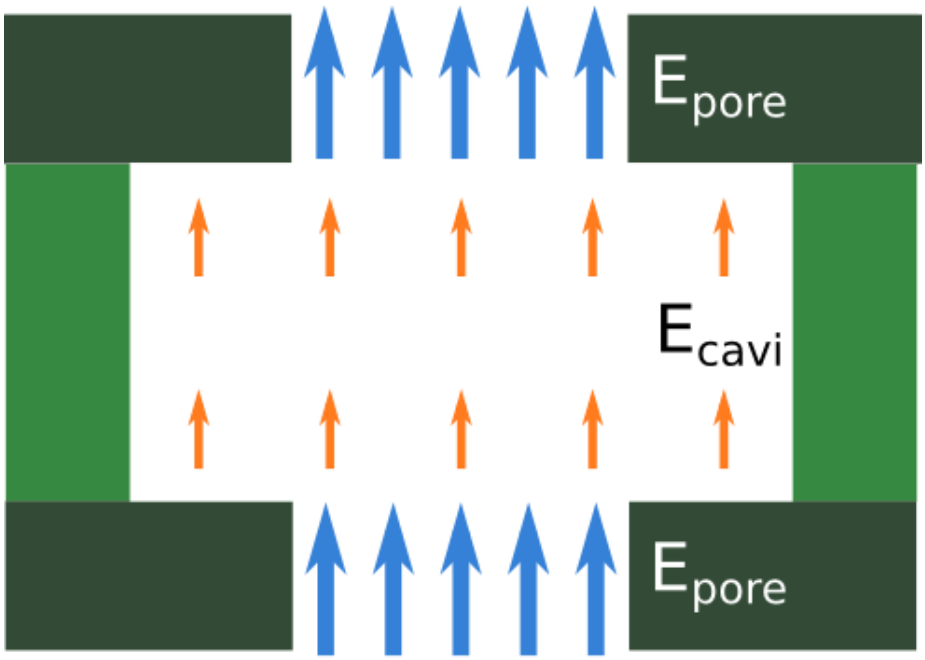

Figure 2.2: Schematic illustrating the electric field **E** in the CN. The blue arrows in the entrance and exit pores are larger and more closely packed than the orange arrows in the cavity, indicating a stronger field. The field is zero outside the CN. Throughout the system, **E** is parallel to the axis of the pore.

is the surface integral of the electric field over that plane,

$$\begin{aligned}\Phi(z) &= \oint (\mathbf{E}\cdot\mathbf{n})\,dS \\ &= \iint ((0,0,E)\cdot(0,0,1))\,dxdy \\ &= \iint E(z)dxdy \\ &= E(z)\iint dxdy,\end{aligned}$$

where the integral is taken over the entire plane. The electric field must be zero

inside the membrane, which is approximated as a perfect insulator. Since both the SN and the CN are cylindrically symmetric, the cross-section at any z is a circle. So the flux at z is

$$\Phi(z) = E(z)\pi r(z)^2, \tag{2.1}$$

where $r(z)$ is the radius of the membrane wall, and so $\pi r(z)^2$ is the cross-sectional area where the field is non-zero.

By conservation of electric flux, the flux must be constant through all the axial segments, i.e. $\Phi(z) = \Phi_0$. From this, we can solve for the electric field as a function of z:

$$E(z) = \frac{\Phi_0}{\pi r(z)^2}. \tag{2.2}$$

Inside the nanopores, the electric field is non-zero until the walls of the membrane. As discussed in the Introduction, the interactions between the walls and the monomers is represented using the WCA potential. This model of the walls corresponds loosely to a hard wall existing at the effective radius r_{eff}. This is where the electric field will be assumed to go to zero. As such, inside the nanopores the electric field is

$$E(z) = \frac{\Phi_0}{\pi r_{\text{eff}}(z)^2}, \tag{2.3}$$

where $r_{\text{eff}}(z)$ is the radius of the cavity in the entrance pore, cavity, or exit pore.

At z values outside the nanopores, the walls of the experiment extend to a distance much greater than the width of the nanopores. This will be modelled

as $r(z) \to \infty$, so that the electric field outside the pore is

$$\lim_{r(z)\to\infty} E(z) = \lim_{r(z)\to\infty} \frac{\Phi_0}{\pi r(z)^2} = 0. \tag{2.4}$$

So the electric field is only non-zero inside the nanopores. The region that qualifies as inside of the CN has a thickness equal to the effective thickness of the membrane, rather than its nominal thickness, for the same reasons that the field is bounded at the effective radius.

In the SN, $r_{\text{eff}}(z)$ is constant throughout the pore. Thus the above equations yield an electric field that is zero outside the pore and that is constant and axial inside the nanopore.

Conversely, in the CN, $r_{\text{eff}}(z)$ varies with z. Since the entrance and exit pores are the same size, the electric field in each of them is the same. This field will be called E_{pore}. The electric field in the cavity, E_{cavi}, will be weaker, given by

$$\frac{E_{\text{cavi}}}{E_{\text{pore}}} = \frac{\Phi_0}{\pi r_{\text{eff,cavi}}^2} \Bigg/ \frac{\Phi_0}{\pi r_{\text{eff,pore}}^2}, \tag{2.5}$$

$$\implies E_{\text{cavi}} = E_{\text{pore}} \left(\frac{r_{\text{eff,pore}}}{r_{\text{eff,cavi}}} \right)^2. \tag{2.6}$$

The units of electric charge in the system are chosen to make each monomer have a charge of $q = 1$, with all the monomers having equal charge, so that the electric force on the monomer at any point is given by $F = qE$. The force experienced by monomers in the pores of the CN will be called F_{pore}, and that experienced by those in the cavity will be called F_{cavi}.

2.1.4 Initial Conditions and Equilibration

The positions of the monomers at $t = 0$ are called the system's initial conditions. The initial conditions for the translocation simulations should correspond to polymers' actual conformations at the beginning of experimental translocation. However, this initial conformation is a complicated function of experimental conditions, so it is common for simulations to use simplified initial conditions. This section explains the initial conditions used for the current study.

The system is initialized with some number of seed monomers already inside the nanopore. One seed monomer was used for the SN case, whereas three seed monomers were used for the CN case. Numerical integration proceeds until no monomers remain inside the pore.

The initial conformation of the polymer is along a straight line through the axis of the nanopore. This is a very thermodynamically inaccessible state, and is unlikely to represent the configuration of any polymer during an experiment. The correct initial conformation of the polymer at the start of translocation is a matter of active research in the cNAB.LAB. However, the most common procedure in computational simulations of nanopore translocation is to fix the seed monomers in space for some equilibration time while the rest of the polymer is allowed to fluctuate under the influence of the LD equations. If the equilibration time is long enough, the resulting conformation is expected to be a fair sample from the equilibrium distribution of polymer conformations subject to the fixation of the seed monomers. Whether sampling the initial configuration of the polymer in this fashion corresponds to physical reality is beyond the scope of this thesis. Similarly, these simulations do not make any predictions concerning the rate at which polymers are captured by the pore from free solution; this is also a matter of active research in the cNAB.LAB.

The rate of randomization of the polymer conformation can be quantified by

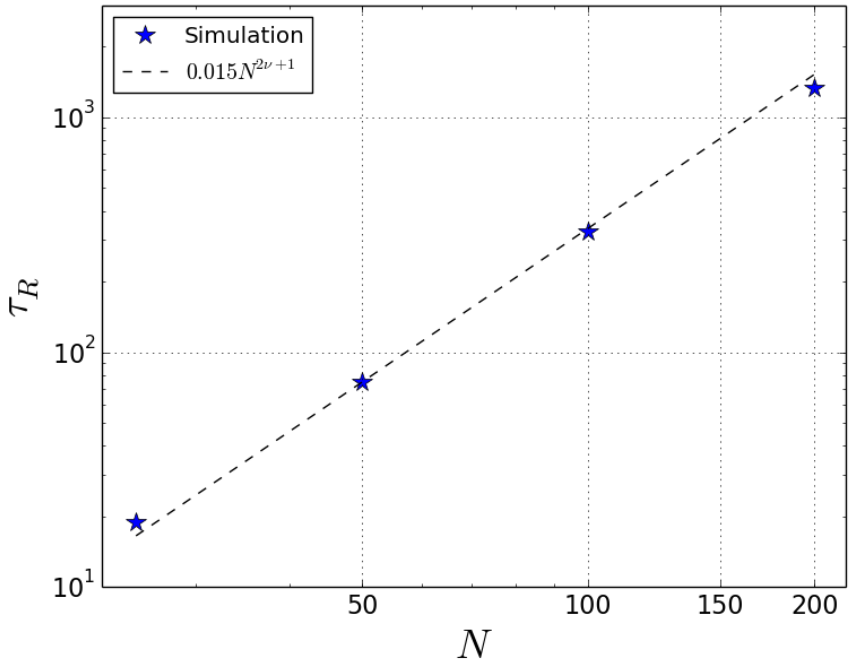

Figure 2.3: Relaxation times τ_R as a function of chain length N as measured in the simulations, compared to the scaling relation predicted by the Rouse polymer model.

the relaxation time τ_R of the polymer, as described in the Introduction. Fig 2.3 demonstrates the result of calculating the relaxation time of various polymer lengths using the current simulation model. The measured relaxation times are compared to the scaling law predicted by the Rouse polymer model, as described in the Introduction. If the equilibration time is set larger than a few multiples of τ_R, the polymer can be expected to be sufficiently equilibrated.

A useful simulation technique for accelerating the equilibration process is to set the friction coefficient artificially low during equilibration. This allows monomers to move more rapidly, so that the polymer can access more conformations more rapidly. As long as the temperature is maintained the same, the equilibrium distribution of conformations won't be changed by this technique.

For these simulations, γ was set to 0.1 during equilibration, compared to 1 during simulations.

2.1.4.1 Failed Events

The simulation procedure outlined above initializes the polymer in a captured state. In other words, it omits the capture process, outlined in Figure 1.5, where the polymer first moves from free solution into the nanopore. As discussed previously, although this does affect the results of the simulation, this is in line with common practices in the literature.

However, although the polymer is initialized in the captured state, it is still possible for it to retract entirely to the *cis* side of the pore. Such an occurrence is referred to as a failed translocation event. Indeed, this behaviour is witnessed in experiments, where failed translocation events typically appear as perturbations in the ionic current that are too brief to be successful translocation events. After a failed event, the electric field outside the pore biases the polymer motion so that it is likely to attempt translocation again thereafter.

However, in the current simulation methodology, the polymer is exceedingly unlikely to re-enter the pore after retraction, since the electric field is set to zero outside the pore. The polymer must overcome a significant entropic barrier in order to thread into the pore from unbiased diffusion in free solution. As a result, if a polymer retracts entirely to *cis* at any point in the simulation, the event is terminated and recorded as a failed translocation event, and the simulation proceeds with a new translocation event. Appendix A shows the rates at which these failed events occur for certain simulation parameters.

This procedure introduces some bias into the results, as it removes translocation events that fail once or more before eventually translocating successfully. Future work will repeat the study in this thesis using an electric field that extends beyond the pore, which will enable a more complete study of the capture

process.

2.1.5 Simulation Details

Table 2.1 shows the parameter values common to all simulations, whereas Table 2.2 summarizes the simulation conditions for which data was collected. Not all combination of conditions were explored. The simulation results are presented in the next sections.

Simulations were conducted for nominal entrance and exit pore radii of 1.3σ and 1.5σ. The simulations with $r_{\mathrm{nom,pore}} = 1.5\sigma$ were conducted first, but it was discovered that these pores occasionally allowed more than one monomer to be in these pores simultaneously. This can lead to so-called hair-pin events, where the polymer can translocate through a pore by a monomer that is not one of its free ends. Hair-pin events present an entire new regime of polymer dynamics, and these were beyond the scope of the current work. Although hair-pin events were not very common at $r_{\mathrm{nom,pore}} = 1.5\sigma$, the simulations were repeated using the smaller pore size to force single-file passage of monomers. Future work will explore the dynamics of hair-pin events in the cavity-nanopore system.

The parameters in Table 2.1 are expressed in terms of fundamental unit scales:

- The mass scale is set by m, the mass of a single monomer.
- The energy scale is set by k_BT, the thermal energy of the system.
- The energy scale of the WCA potential is set to $\epsilon = k_BT$.
- From this, the length scale is determined by the distance σ at which the WCA potential energy is equal to $\epsilon = k_BT$.
- In this sense, the length σ corresponds to the diameter of the monomers.

Parameter	Value
Monomer mass	$m = 1$
Temperature	$k_B T = 1$
WCA energy scale	$\epsilon = 1 k_B T$
WCA length scale	$\sigma = 1$
WCA cut-off distance	$r_c = \sqrt[6]{2}\sigma$
FENE spring constant	$\kappa = 30\epsilon/\sigma^2$
FENE maximum length	$r_{\max} = 1.5\sigma$
Friction coefficient	$\gamma = 1\sqrt{\frac{m\epsilon}{\sigma^2}}$
Timestep size	$\Delta t = 0.01\sqrt{\sigma^2 \frac{m}{\epsilon}}$
Equilibration time	$10^3\sqrt{\sigma^2 \frac{m}{\epsilon}} = 10^5 \Delta t$
SN pore thickness	$t_{\text{eff}} = 1.0001\sigma$
CN entrance/exit pore thickness	$t_{\text{eff}} = 1.0001\sigma$
SN seed monomers	1
CN seed monomers	3

Table 2.1: Simulation parameters that are common to all simulation scenarios.

All other quantities are expressed in terms of these three units, m, ϵ, σ, as derived below.

The units of the FENE spring constant can be derived directly from the form of the FENE potential. Starting with Equation 1.42,

$$U_{\text{FENE}}(r) = -\frac{1}{2}\kappa r_{\max}^2 \ln\left(1 - \frac{r^2}{r_{\max}^2}\right), \tag{2.7}$$

taking the units of each side yields

$$\epsilon = [\kappa]\sigma^2 \implies [\kappa] = \frac{\epsilon}{\sigma^2}, \tag{2.8}$$

where $[\cdot]$ is used to indicate the dimensions of a quantity. Note that the input and output of elementary functions, like $\ln(\cdot)$, are always dimensionless.

The units of the friction coefficient γ are obtained using Equation 1.33 for

the average motion of a free monomer. Starting with

$$m \langle \mathbf{a} \rangle_t = -\gamma \langle \mathbf{v} \rangle_t + \mathbf{F}_{\text{ext}}, \tag{2.9}$$

taking the units of each side yields

$$m\frac{\sigma}{]\tau]^2} = [\gamma]\frac{\sigma}{[\tau]} \implies [\gamma] = \frac{m}{[\tau]}, \tag{2.10}$$

where $[\tau]$ is the unit of time. In the current formalism, however, the time scale can be expressed in terms of m, ϵ, σ. Consider for instance Equation 1.30 for the MSD in the long-time limit,

$$\lim_{t\to\infty} \left\langle (\mathbf{r}(t) - \mathbf{r}(0))^2 \right\rangle = 6Dt \tag{2.11}$$

$$= 6\frac{k_B T}{\gamma} t. \tag{2.12}$$

Taking the units yields

$$\sigma^2 = \frac{\epsilon}{[\gamma]}[\tau] \implies [\gamma] = \frac{\epsilon}{\sigma^2}[\tau] \tag{2.13}$$

Combining Equations 2.10 and 2.13 yields

$$\frac{m}{[\tau]} = \frac{\epsilon[\tau]}{\sigma^2} \implies [\tau] = \sqrt{\sigma^2 \frac{m}{\epsilon}}. \tag{2.14}$$

Finally, using Equation 2.14 with Equation 2.10 yields the units for γ,

$$[\gamma] = \frac{m}{[\tau]} = \frac{m}{\sqrt{\sigma^2 \frac{m}{\epsilon}}} = \sqrt{\frac{m^2}{\sigma^2 \frac{m}{\epsilon}}} = \sqrt{\frac{m\epsilon}{\sigma^2}} \tag{2.15}$$

The plots given in the next sections do not include error bars. This is because sufficient data was acquired that the error bars would be smaller than

Parameter	Values
Polymer Chain Length	$10 < N < 400$
SN pore size	$r_{\mathrm{nom,pore}} = 1.3\sigma, 1.5\sigma$
CN entrance/exit pore size	$r_{\mathrm{nom,pore}} = 1.3\sigma, 1.5\sigma$
CN cavity size	$r_{\mathrm{eff,cavi}} = 3.0\sigma, 3.5\sigma, 4.0\sigma$
Force	$0.4 < F_{\mathrm{pore}} < 4.0$

Table 2.2: List of the simulations parameters that were varied and the ranges that were explored.

the size of the markers used to represent the data. Since the size of the error bars depends on the intrinsic standard deviation on a given data point, the number of samples required to make the error bars this small depended on the simulation conditions. Overall, each data point represents the average of several hundred simulations using distinct random seeds.

2.2 Cavity-Nanopore Results in the Basic Regime

This section presents the results of simulations used to characterize the dynamics of polymer translocation through the cavity-nanopore (CN). First, the results in the low-F_{pore} regime with narrow entrance/exit pores and a large cavity will be presented, as they illustrate the novel dynamics of the system most clearly. These conditions will be referred to as the basic regime of CN operation. The most striking result is that the translocation time is a complicated non-monotonic function of chain length. Specifically, translocation is fastest for chains of intermediate length: both shorter and longer chains take longer to translocate.

In order to explore this behaviour, the translocation process is subdivided into three phases. Analysis of the polymer dynamics during each of these phases is used to explain the non-monotonic dependence of translocation time on chain length. Parallels are drawn to simulations of translocation through a standard nanopore (SN). The qualitative picture of CN translocation that emerges from

this analysis is discussed in terms of a free energy landscape.

After completing the discussion of translocation dynamics in the basic regime, results are presented for modified translocation conditions. First, the effects of varying the cavity size are presented, followed by the effects of varying the magnitude of the force field, and then finally by the effects of widening the entrance/exit pores. In all cases, the results are discussed as deviations from the dynamics observed in the basic regime.

Next, using the qualitative description of the translocation dynamics in the basic regime of CN operation, a free energy model is established to predict the chain length for which translocation is fastest. The predictions of the model are compared against simulation results across different cavity sizes, force intensities, and entrance/exit pore sizes. Good agreement is recovered near the basic regime, and disagreements in other regimes are explained.

Finally, possible applications of the CN nanopore are discussed. Ongoing work and future research directions are summarized.

2.2.1 Translocation Time

Figure 2.5 shows the average translocation time through the CN t_{trans} as a function of polymer length N when $F_{\text{pore}} = 0.4\epsilon/\sigma, r_{\text{nom,pore}} = 1.3\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. For comparison, Figure 2.4 shows the average translocation time through an equivalent SN with $F_{\text{pore}} = 0.4\epsilon/\sigma$ and $r_{\text{nom,pore}} = 1.3\sigma$. Both plots are shown on log-log scales.

The SN results in Figure 2.4 are fit to a power law relationship of the form

$$\tau \propto N^{\alpha}. \tag{2.16}$$

The data obtained here fit well for $\alpha \approx 1.47$, which is in line with the range of exponents published in the literature, as described in the Introduction. This

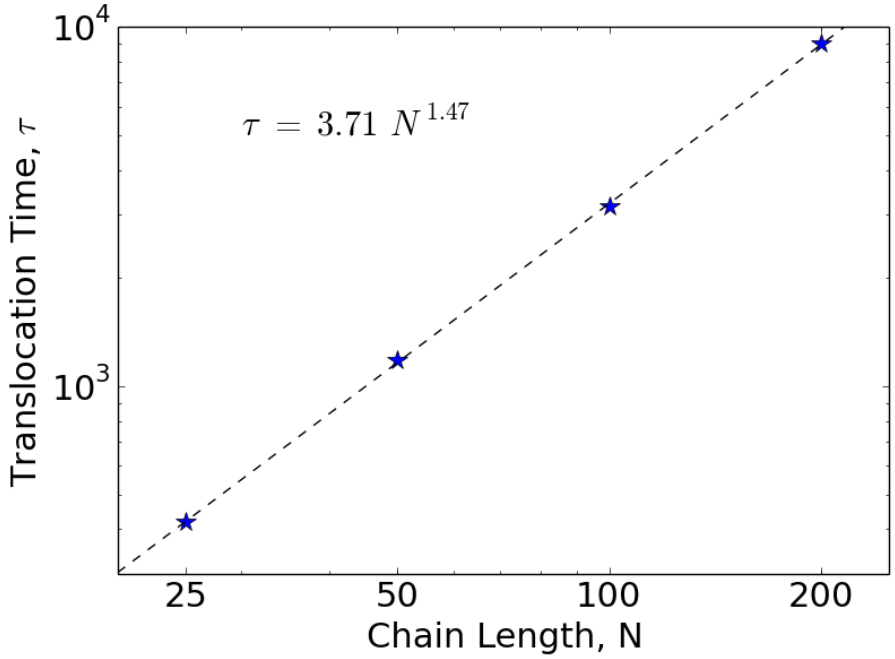

Figure 2.4: Translocation times through the standard nanopore for $F_{\text{pore}} = 0.4\epsilon/\sigma$ when $r_{\text{nom,pore}} = 1.3\sigma$. The power law $\tau = 3.71N^{1.47}$ provide a good fit to the data.

lends credence to the current simulation methodology and implementation. Note that power law relationships appear as straight lines on log-log plots, because

$$\tau \propto N^{\alpha} \implies \log(\tau) \propto \alpha \log(N). \tag{2.17}$$

The CN results in Figure 2.5 evidently cannot be fit to a single power law relationship. The behaviour of t_{trans} as a function of N can be described in three regimes. For short chains, t_{trans} decreases monotonically with N. This continues until $N \approx 75$, whereafter t_{trans} increases very abruptly with N until $N \approx 95$. For chains longer than this, t_{trans} remains roughly constant with N, perhaps growing slowly.

Figure 2.5 demonstrates the single most striking feature of CN translocation:

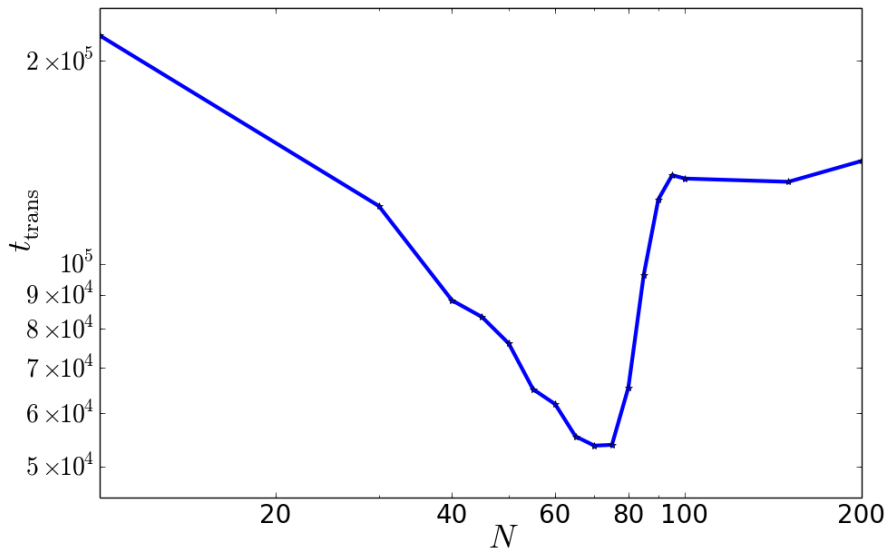

Figure 2.5: Translocation times through the cavity-nanopore for $F_{\text{pore}} = 0.4\epsilon/\sigma$ when $r_{\text{nom,pore}} = 1.3\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. Note the pronounced minimum at $N \approx 75$.

t_{trans} is a non-monotonic function of N with a very sharp minimum at intermediate N values. That is, medium-length chains translocate more quickly than both shorter and longer chains. This is in stark contrast with the SN results in Figure 2.4, where longer chains always have longer average translocation times. The chain length at which translocation time is minimual will be referred to as the critical chain length N^*.

2.2.2 Three Stages of Translocation

In order to explain the novel translocation dynamics of Figure 2.5, the translocation process will be subdivided into three phases. The state of the polymer at any moment in time will be characterized by three numbers: N_{cis}, the number of particles on the initial side of the membrane (called the *cis* side); N_{cavi}, the number of particles inside the cavity; and N_{trans}, the number of monomers on the final side of the membrane (called the *trans* side). Figure 2.6 shows a plot

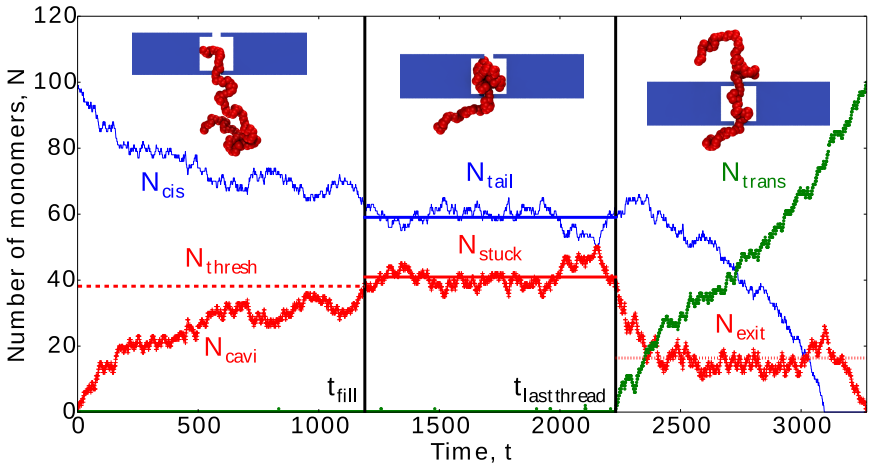

Figure 2.6: Example of the three stages of translocation through the CN. The blue line shows $N_{\rm cis}$, the red crosses show $N_{\rm cavi}$, and the green dots show $N_{\rm trans}$. The first black vertical lines shows $t_{\rm fill}$, marking the end of the fill phase and the beginning of the stuck phase. The second black vertical line shows $t_{\rm lastthread}$, marking the end of the stuck phase and the beginning of the exit phase. The graphics illustrate the conformation of the polymer during each phase. The quantities $N_{\rm thresh}, N_{\rm tail}, N_{\rm stuck}$, and $N_{\rm exit}$ are noted with horizontal lines.

of $(N_{\rm cis}, N_{\rm cavi}, N_{\rm trans})$ over the course of an example translocation event.

As described in the last section, the polymer begins the translocation process with its three leading monomers threaded through the entrance pore and the rest of the polymer equilibrated in this *cis* region. From there, the chain begins to enter the cavity one monomer at a time. The monomers already inside the cavity produce a pressure against the introduction of new particles into the cavity. Eventually, this pressure will balance the electric force, and no additional particles will enter the cavity on average. This first part of the translocation process will be called the filling phase. Its duration will be called $t_{\rm fill}$.

The filling phase ends when the occupancy of the cavity reaches the quasi-equilibrium value alluded to in the previous paragraph. However, this value is not known *a priori*. Instead, the filling phase is considered to be complete when the cavity occupancy $N_{\rm cavi}$ exceeds for the first time a threshold value $N_{\rm thresh}$.

This threshold is estimated as follows. For long chains, the entire chain will never be found in the cavity simultaneously. In this case, the particles in the cavity are treated as a smaller polymer. The threshold N_{thresh} is set to be the length at which this smaller polymer has a radius of gyration R_G equal to the effective radius of the cavity. Using $R_G = C_G N^{\nu}$ for freely-jointed chains with excluded volume, this yields

$$r_{\text{eff,cavi}} = R_G = C_G N_{\text{thresh}}^{\nu}, \tag{2.18}$$

$$N_{\text{thresh}} = \left(r_{\text{eff,cavi}}/C_G\right)^{1/\nu}. \tag{2.19}$$

For the threshold, the accepted value of $\nu \approx 0.588$ was used, and $C_G = 1/\sqrt{6}$, the coefficient for ideal chains, was used. This value of C_G is not correct, but is an acceptable approximation, as the results of the analysis are not very sensitive to the exact choice of threshold.

For small chains, this formula will yield $N_{\text{thresh}} > N$, and the chain will never be considered to have filled the cavity. This is expected, as sufficiently small chains can fit entirely inside the cavity. However, it is natural to consider the filling phase for such a small chain to be complete once it is entirely located inside the cavity. Indeed, for these cases, the choice of threshold is $N_{\text{thresh}} = N$.

The filling phase is only well-defined if the polymer comes to a quasi-equilibrium state in which the leading end of the polymer is still trapped inside the cavity until after $N_{\text{cavi}} > N_{\text{thresh}}$ for the first time. Conversely, it is possible for the leading monomers to immediately thread through the exit pore, so that particles never accumulate inside the cavity. These cases are referred to as "straight-through" events. They can be identified quite easily, for instance by noting that $t_{\text{fill}} = t_{\text{trans}}$ for straight-through cases. Analysis of the simulations reveals that straight-through events amount to fewer than 1% of cases in all simulation scenarios considered in this thesis, so they are simply removed from

the data and neglected for this study.

The next part of the translocation process entails the metastable state wherein N_{cavi} oscillates about the quasi-equilibrium value described above, and the polymer does not yet pass through the exit pore onto the *trans* side of the membrane. This is called the stuck phase. The behaviour of the polymer during this phase turns out to be essential to the overall translocation dynamics. The following quantities are of interest:

- t_{stuck}: The duration of the stuck phase.
- N_{stuck}: The time-averaged number of particles in the cavity during the stuck phase.
- N_{tail}: The time-averaged number of particles on the *cis* side during the stuck phase.

During the stuck phase, one or both ends of the polymer will enter the exit pore and begin threading towards the *trans* side (due to the action of the electric field). These threading attempts will often fail, as threading is very entropically disfavourable. Eventually, however, the polymer will successfully thread through the exit pore and begin exiting the cavity. The time when this occurs is referred to as $t_{\mathrm{lastthread}}$, and marks the end of the stuck phase and the beginning of the final phase. As this phase is dominated by the exit of the polymer from the cavity onto the *trans* side of the membrane, it is called the exit phase. The duration of the exit phase is called t_{exit}, and the average number of monomers present in the cavity during the exit phase is called N_{exit}.

2.2.2.1 Fill Phase

The filling phase was not explored in depth during this study. For all of the simulation conditions considered, t_{fill} is a negligible fraction of t_{trans}. For short chains, $t_{\mathrm{stuck}} \gg t_{\mathrm{fill}}$, whereas for long chains $t_{\mathrm{exit}} \gg t_{\mathrm{fill}}$.

Although the fill phase is generally a small fraction of the total translocation time, the dynamics of the polymer during the fill phase will influence the conformation of the polymer at the beginning of the stuck phase. For this reason, the fill phase will be the subject of further analysis in future work.

2.2.2.2 Stuck Phase

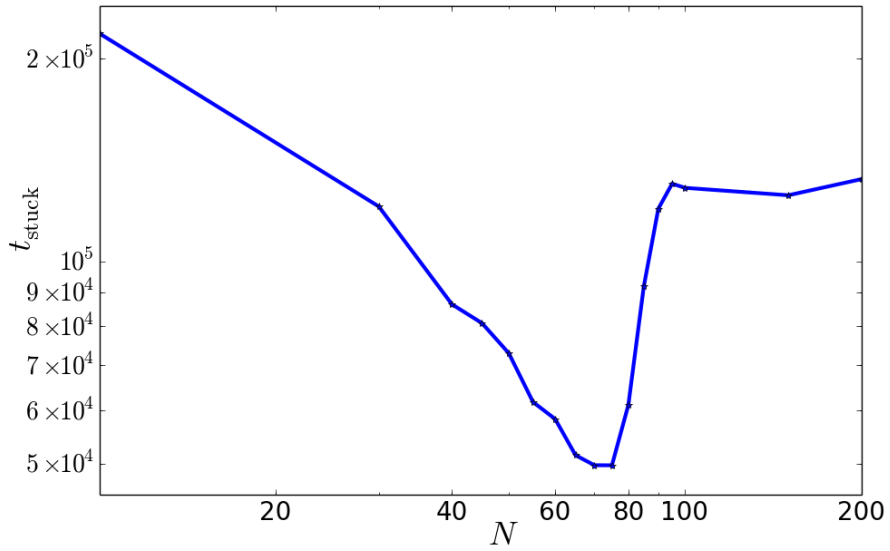

Figure 2.7: Average duration of the stuck phase, t_{stuck}, for the cavity-nanopore with $F_{\text{pore}} = 0.4\epsilon/\sigma, r_{\text{nom,pore}} = 1.3\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. The axes are the same as in Figure 2.5 to emphasize the similarity between the two plots.

Figure 2.7 shows the duration of the stuck phase, t_{stuck}, as a function of chain length N for the basic regime. By comparing Figure 2.5 to Figure 2.7, it is evident that in the basic regime the stuck phase is by far the longest portion of the translocation process, and in fact $t_{\text{trans}} \approx t_{\text{stuck}}$.

Given this, the previous discussion of t_{trans} applies identically to t_{stuck}. In particular, this means that the dynamics that lead to the sharp minimum in t_{trans} must occur during the stuck phase. Since this minimum is the result of primary interest, this calls for a detailed analysis of the polymer behaviour

during the stuck phase.

Throughout the stuck phase, $N_{\text{trans}} \approx 0$: the monomers are either inside the cavity or in the *cis* region. This motivates the two metrics, shown in Figure 2.6, used to characterize the polymer's state during the stuck phase. The first, N_{stuck}, is the average number of monomers inside the cavity during the stuck phase. The second, N_{tail}, is the average number of monomers in the *cis* region during the stuck phase. Since $N_{\text{trans}} \approx 0$, we must have $N \approx N_{\text{stuck}} + N_{\text{tail}}$; however, it is still fruitful to look at plots of N_{stuck} and N_{tail} separately.

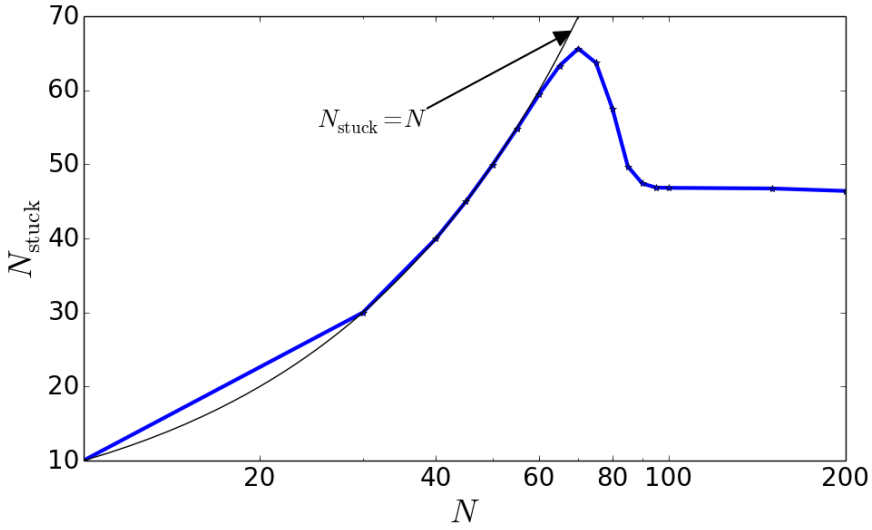

Figure 2.8: Plot of N_{stuck}, the average of N_{cavi} during the stuck phase, for the cavity-nanopore with $F_{\text{pore}} = 0.4\epsilon/\sigma, r_{\text{nom,pore}} = 1.3\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. The black line shows $N_{\text{stuck}} = N$ for reference.

Figures 2.8 and 2.9 show plots of N_{stuck} and N_{tail}, respectively, for the CN operating in the basic regime. For short chains, the entire chain fits inside the cavity, so $N_{\text{stuck}} \approx N$ and $N_{\text{tail}} \approx 0$. Above $N \approx 70$, chains begin to form a tail: N_{tail} starts to grow with N. Note that this is roughly the same range of chain lengths that minimizes t_{stuck} and t_{trans}. This suggests that the presence of a tail during the stuck phase somehow corresponds to the sharp rise in t_{stuck}

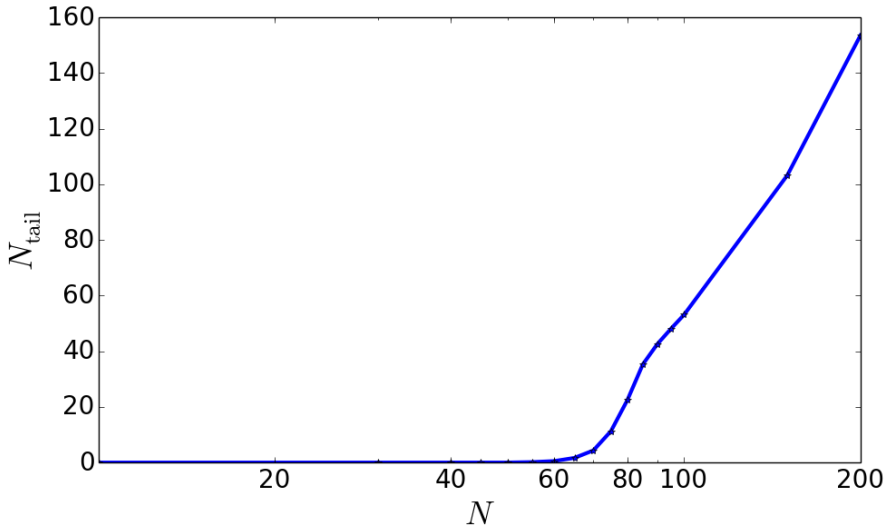

Figure 2.9: Plot of N_{tail}, the average of N_{cis} during the stuck phase, for the cavity-nanopore with $F_{\text{pore}} = 0.4\epsilon/\sigma, r_{\text{nom,pore}} = 1.3\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$.

and t_{trans} for $N > N^*$.

2.2.2.2.1 Stuck Phase as an Escape Process In order to understand this behaviour, we will treat the stuck phase as a quasi-equilibrium process. That is, although the stuck phase is not the final state of the translocation process, it is a metastable state that occurs during translocation and that lasts a long time. In particular, comparing to the measured relaxation times shown in Figure 2.3, the duration of the stuck phase is at least two orders of magnitude larger than the relaxation times of the polymers in free solution, i.e. $t_{\text{stuck}} \gg \tau_R$. As such, it is plausible that the polymer comes to a quasi-equilibrium state during the stuck phase.

With this perspective, the transition from the stuck phase to the exit phase can be described as an escape process. The polymer conformations during the stuck phase have relatively many degrees of freedom: the monomers in the cavity are free to explore the cavity volume. Conversely, threading through the exit

pore entails fixing one of the free ends of the chain to that pore. This restriction in the degrees of freedom of the polymer correspond to the entropic barrier to escape. Thus t_{stuck} can be thought of as the time required for the polymer overcome this entropic barrier via thermal fluctuations. Escape is encouraged via the action of the electric field in the exit pore, but this field can only affect monomers inside the pore. Escape processes generally depend exponentially on the height of the free energy barrier, explaining why t_{stuck} is so large.

Having established t_{stuck} as the time required to escape across a free energy barrier, the effect of N on t_{stuck} can be understood by its effect on the height of that barrier. The barrier is roughly the difference between the free energy of the chain in the stuck state and the free energy of the chain at the beginning of the exit phase. The following discussion will consider how this barrier changes with increasing N, and elucidate the role of the tail in the rapid increase of t_{stuck}.

Consider first a chain that is short enough to fit inside the cavity with hardly any deformation. The entropic barrier is large, as the chain is barely restricted in the stuck state, but will be greatly restricted in the exit state.

As N is increased, the chain becomes progressively more confined inside the cavity during the stuck phase. The free energy cost of remaining in the stuck phase grows more rapidly with N than the free energy cost of initializing the exit phase. Thus, the free energy barrier decreases, and t_{stuck} decreases with N, as seen in Figure 2.7 for $N \lesssim 70$.

Increasing N beyond this point leads to the formation of tails during the stuck phase. Whereas $N_{\text{stuck}} \approx N$ for shorter chains, the introduction of a tail allows N_{stuck} to drop well below N. In fact, as N_{tail} first becomes non-zero, N_{stuck} decreases significantly. This occurs because, in the presence of a tail, the polymer can regress partially back to the *cis* side during the stuck phase. For short chains regression through the entrance entails re-threading through

the entrance pore in addition to moving against the electric field in the pore. Conversely, the presence of a tail during the stuck phase means that the entropic cost of re-threading is absent. The energetic barrier imposed by the electric field can be overcome by thermal fluctuations. Thus, as soon as N is large enough that the polymer cannot fit entirely in the cavity during the stuck phase, N_{stuck} drops to much lower values.

Now consider the height of the free energy barrier to exit for a polymer that has a tail during the stuck phase. Since the presence of the tail has decreased N_{stuck}, the free energy cost of confinement in the cavity has been decreased. Meanwhile, the free energy cost of being in the exit phase has not been decreased. Thus, the free energy barrier to escape has been increased, and t_{stuck} increases.

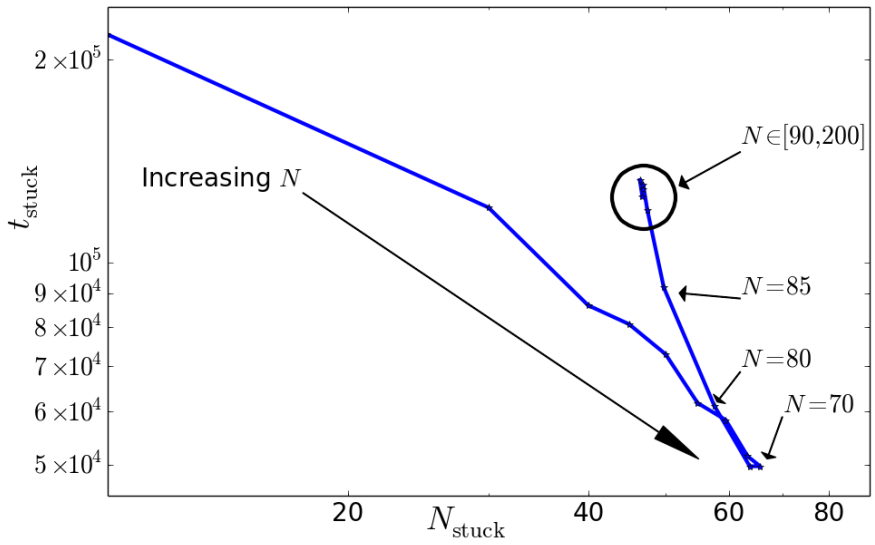

Figure 2.10: Plot of t_{stuck} against N_{stuck} for the cavity-nanopore with $F_{\text{pore}} = 0.4\epsilon/\sigma, r_{\text{nom,pore}} = 1.3\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. The labels indicate the order of increasing N.

This discussion proposes that the dominant factor in determining the free energy barrier to escape is the occupancy of the cavity, N_{stuck}. This would suggest that t_{stuck} is also determined predominantly by N_{stuck}. To test this

hypothesis, Figure 2.10 shows t_{stuck} plotted against N_{stuck} for $N \in [10, 200]$. For $N \in [10, 70]$, this shows t_{stuck} decreasing monotonically with N_{stuck}. This corresponds to the tail-less regime, where $N_{\text{stuck}} \approx N$. After $N > 70$, this trend reverses itself.

For $N \in [70, 80]$ however, t_{stuck} remains a single-valued function of N_{stuck}: in this regime, where the tail has formed but is short, the stuck time is uniquely determined by N_{stuck}. Increasing N even more causes t_{stuck} to deviate above its previous values. This suggests that sufficiently long tails begin to directly affect the transition between the stuck and exit phases. The rapid increase in t_{stuck} after N^* corresponds roughly to the range $N \in [70, 95]$. Thus, this phenomenon is due both to the rapid decrease in N_{stuck} caused by the tail enabling regression to *cis* as well as the increase in $t_{\text{stuck}}(N_{\text{stuck}})$ due to the tail affecting the escape dynamics directly.

In summary, the stuck phase is a quasi-equilibrium process where the polymer tries to escape from the metastable stuck state to the exit state over a free energy barrier. For short chains, increasing N reduces the barrier height, and thus decreases t_{stuck}. When N is large enough, a tail forms in the *cis* region. As soon as the tail is present, regression from the cavity to the *cis* region during the stuck phase is possible, so N_{stuck} decreases rapidly with N. The decrease in N_{stuck} increases the height of the barrier to escape, so there is a concomitant rapid increase in t_{stuck}. Furthermore, the dynamics of the tail directly affect the escape process, so that t_{stuck} at a given value of N_{stuck} is larger when a tail is present. These two mechanisms combine to create the dramatic increase in t_{stuck} for $N > N^*$.

2.2.2.3 Exiting Phase

In the basic regime, $t_{\text{exit}} \ll t_{\text{stuck}}$, so the exit phase is less important than the stuck phase to the overall translocation dynamics. However, the dynamics

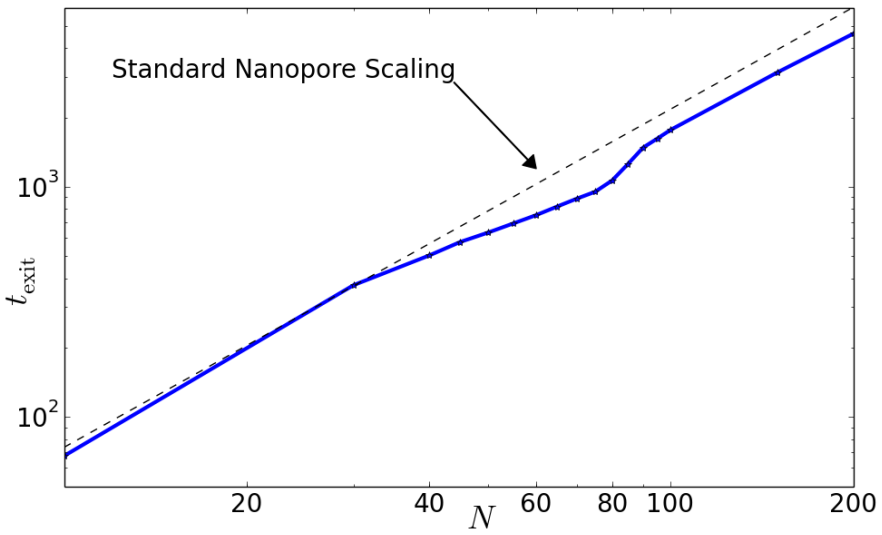

Figure 2.11: The blue line is the duration of the exit phase, t_{exit}, as a function of chain length N for the cavity-nanopore with $F_{\mathrm{pore}} = 0.4\epsilon/\sigma, r_{\mathrm{nom,pore}} = 1.3\sigma$ and $r_{\mathrm{eff,cavi}} = 4.0\sigma$. The dotted line is proportional to $N^{1.47}$, the power law obtained for the translocation time τ through a standard nanopore with $F_{\mathrm{pore}} = 0.4\epsilon/\sigma$ and $r_{\mathrm{nom,pore}} = 1.3\sigma$. The dotted line has been rescaled to line up with the blue line, to facilitate comparison.

of t_{exit} as a function of N are quite interesting from the perspective of basic polymer physics. Figure 2.11 shows the exit time t_{exit} as a function of chain length N, along with a dotted line indicating the scaling of the translocation time τ through a SN equivalent to the exit pore. The t_{exit} is parallel to the SN scaling law for short and very long chains, but deviates from it in between.

Figure 2.12 shows the average number of monomers in the cavity during the exit, N_{exit}, first introduced in Figure 2.6. This metric will yield some light on the conformation of the polymer during the exit process. The shape of N_{exit} is reminiscent of N_{stuck}, which was shown in Figure 2.8. The dotted line in Figure 2.12 compares N_{exit} to $N/2$.

For short chains, the average occupancy of the cavity over the exit process is roughly $N/2$. These chains fit entirely inside the cavity during the stuck phase,

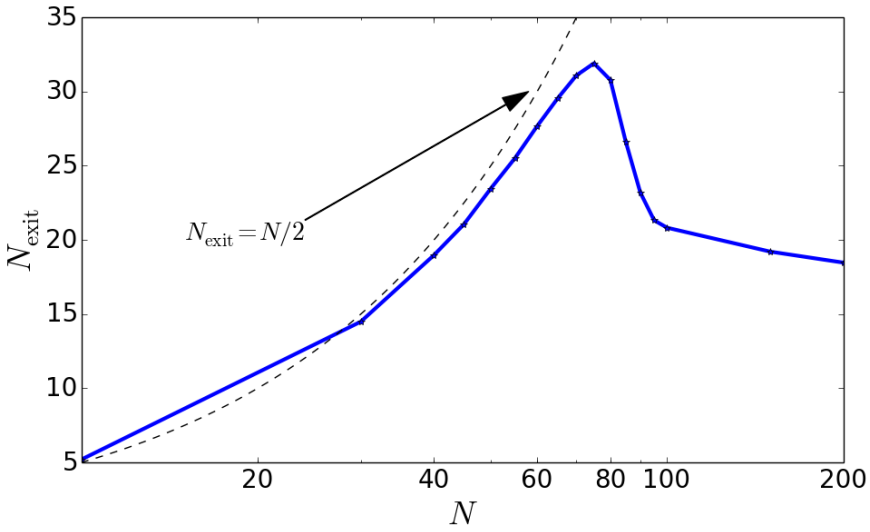

Figure 2.12: The blue line is the occupancy of the cavity, t_{exit}, as a function of chain length N for the cavity-nanopore with $F_{\text{pore}} = 0.4\epsilon/\sigma, r_{\text{nom,pore}} = 1.3\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. The dotted line is equal to $N/2$.

so $N_{\text{cavi}} \approx N$ at the beginning of the exit phase. By definition, $N_{\text{cavi}} = 0$ at the end of the exit phase. So if $N_{\text{exit}} = N/2$, this suggests that the speed of the exit process is roughly constant over the exit phase.

When N becomes large enough to form a tail during the stuck phase, N_{exit} drops rapidly. This corresponds to a sudden change in the initial conformation of the polymer at the onset of the exit phase. When no tail is present, the chain begins the exit phase entirely confined within the cavity. Conversely, when a tail is present, the portion of the chain inside the cavity must connect the entrance and exit pores.

For longer chains, N_{exit} continues to drop, but seems to converge slowly to some limiting value. This value is presumably the equilibrium occupancy of the cavity when the polymer is extended between the entrance and exit pores. Thus the result suggests that long chains spend the majority of the exit phase in this elongated conformation.

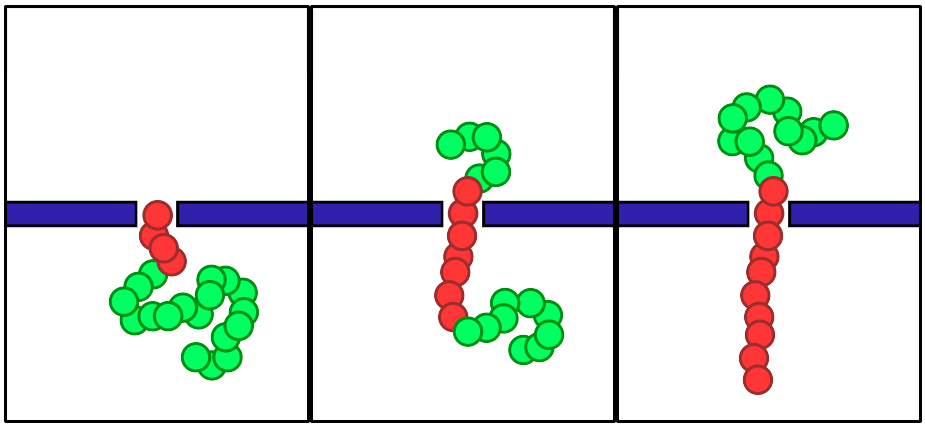

Figure 2.13: Schematic illustrating the tension propagation model for the standard simulation protocol of SN translocation. Green monomers correspond to parts of the chain that are relaxed, whereas red monomers are under tension. The three panels represent the early, middle, and late stages of translocation, left to right respectively.

2.2.2.3.1 Tension Propagation in the Exit Phase The dynamics of the exit phase can be well understood in terms of the tension propagation model for translocation through standard nanopores. The tension propagation model for polymer translocation describes how the force pulling on the monomer in the nanopore propagates as a tension front along the polymer chain. Figure 2.13 illustrates tension propagation in a typical SN translocation simulation. At the beginning of the simulation, the polymer is in a relaxed conformation. As the leading end of the chain is pulled into the pore, the pulling force is transmitted along the chain to adjacent monomers by tension in the polymer bonds. As translocation proceeds, progressively more of the remaining monomers on the *cis* side are under tension.

This tension propagation picture predicts that translocation slows down as the tension front encompasses more monomers. The field in the nanopore acts only on the monomer in the nanopore. Conversely, all of the monomers under tension experience a drag force. Thus, the applied force is constant throughout

the translocation process, but the opposing drag force increases as the tension front propagates along the chain. The dynamics of this tension propagation leads to the scaling of the total translocation time τ as a function of chain length N.

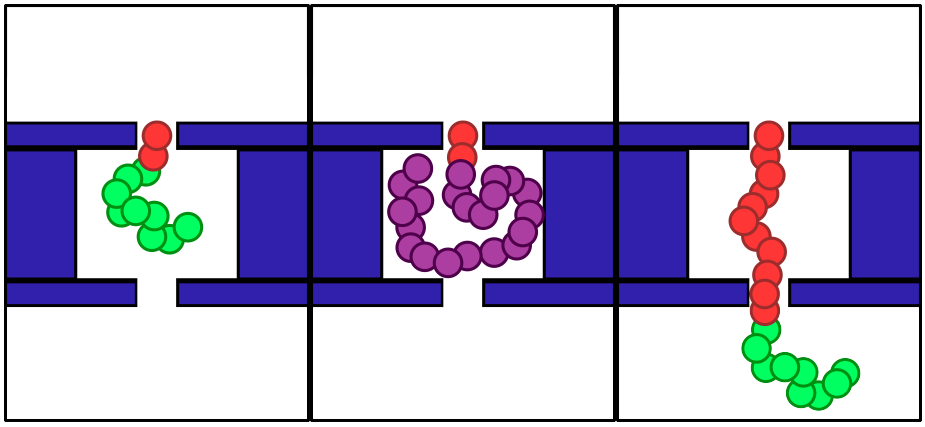

Figure 2.14: Schematic illustrating typical initial conformations of polymers at the onset of the exit phase for different polymer lengths. Green monomers correspond to part of the chain that are relaxes and red monomers are under tension, as in Figure 2.13. Purple monomers indicate parts of the chain that are compressed due to the confinement of the cavity.

Figure 2.14 illustrates typical initial conformations of polymers at the onset of the exit phase for different polymer lengths. The force inside the cavity is small enough in the basic regime that it can be neglected. Sufficiently short chains are barely confined by the cavity. This corresponds to the leftmost panel in Figure 2.14. For these chains, the exit phase is almost identical to translocation through a SN equivalent to the exit pore. Indeed, for short chains t_{exit} in Figure 2.11 agrees with the SN power law.

As N increases, the initial conformation of the polymer at the onset of the exit phase is progressively more confined by the cavity. This is the middle schematic in Figure 2.14. The entropic cost of confinement creates an entropic force pushing the polymer to exit faster. Furthermore, as illustrated in Fig-

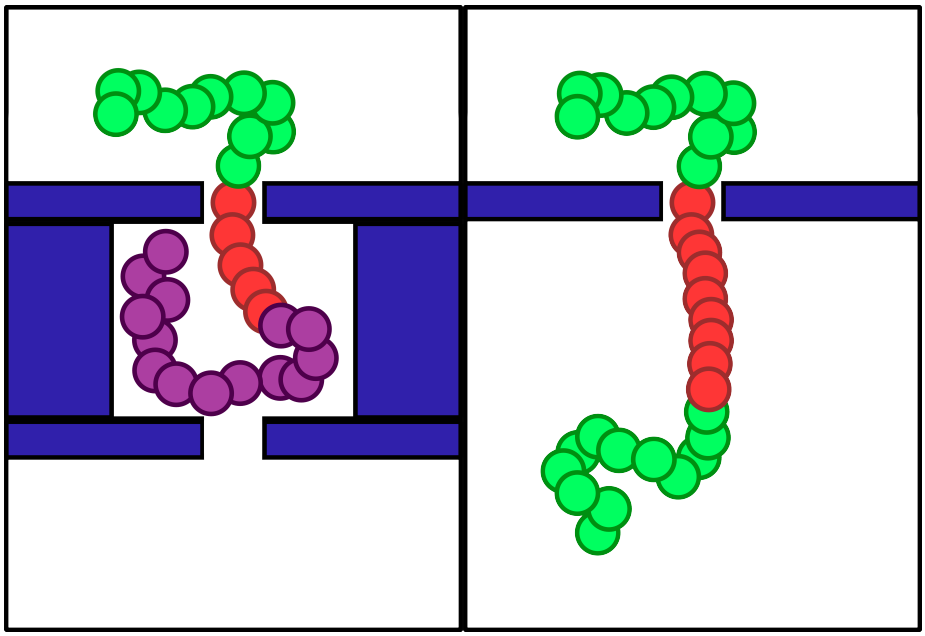

Figure 2.15: Schematic illustrating the difference between a compressed polymer in the exit phase and the translocation of the same polymer without confinement. As in Figure 2.14, red monomers are under tension, green monomers are relaxed, and purple monomers are compressed. The free polymer can form a much longer tension front than the confined polymer.

ure 2.15, the confinement of the cavity changes the effect of tension propagation in the polymer. The cavity ensures that all monomers cannot be farther from the exit pore than the *cis* wall of the cavity. This restricts the maximum number of monomers that can be under tension at a time. Since having more monomers under tension slows down translocation, this means that exit from the cavity is relatively faster than translocation through a SN. This is manifested in t_{exit} as shown in the part of Figure 2.11 where the rate of increase of t_{exit} with N is slower than the rate of increase of τ through a SN for the same polymer length.

This behaviour is akin to results found by Sean et al., who used LD simulations similar to those in this thesis to study polymer translocation through a nanopore where the polymer is initially confined inside a tube on the *cis* side

of the membrane [66]. The confining tube in that work is analogous to the confinement provided by the CN cavity during the stuck phase in the present thesis. Sean et al. also successfully described the dynamics of that system using the tension-propagation model, in an argument similar to that provided in the previous paragraph. Since the study of translocation from a confining tube yielded other interesting dynamics, further studies on the CN will investigate whether any of those results can also be uncovered in the CN system.

The above description of the exit process holds until N is large enough that tails begin to form during the stuck phase. The emergence of tails rapidly changes the initial conformation of the polymer at the onset of the exit phase from the middle schematic in Figure 2.14 to the rightmost schematic. The mechanism described in the previous paragraph that facilitated exit from the cavity no longer occur when a tail is present. Instead, the polymer begins the exit phase in a relatively extended conformation between the entrance and exit pores. The tension front propagates very rapidly between these two points along the chain, making translocation slower than it would be for the same polymer translocating through a SN. Indeed, this corresponds to the region of Figure 2.11 where t_{exit} grows more rapidly than the SN power law.

For sufficiently long chains, the exit process should be dominated by the translocation of the tail. The tension front will rapidly propagate through the cavity into the *cis* region, and the portion of the chain inside the cavity will represent only a small fraction of the total tension front. In this limit, the exit phase should resemble translocation through a normal SN, as the dynamics of the tail in the *cis* region become more important than the dynamics inside the cavity. Indeed, Figure 2.11 show that at large N the exit time t_{exit} resumes the scaling of the SN power law.

In summary, in the basic regime the exit phase dynamics can be described in

four regimes of N. For very short chains, the exit phase is virtually identical to SN translocation. For longer chains that can fit entirely inside the cavity during the stuck phase, the polymer is compressed at the onset of the exit phase. As a result, increasing the length of the chain does not increase the exit time as rapidly as in normal SN translocation: thus t_{exit} grows less rapidly than the SN power law for τ. When the chain is long enough that it forms a tail during the stuck phase, it is no longer as compressed, and is instead elongated at the onset of the exit phase. The elongated polymer translocates more slowly, so t_{exit} grows more rapidly with N than the SN power law for τ. Finally, in the large-N limit, the exit phase is dominated by the dynamics of the tail in the *cis* region. The dynamics inside the cavity become negligble, and t_{exit} resumes a scaling law proportional to the SN power law for τ.

2.2.2.4 Straight-Through Events

The previous discussion postulates that translocation events can be broken down into a filling phase, a stuck phase, and an exit phase. In particular, the stuck phase begins when the cavity occupancy, N_{cavi}, exceeds a threshold value N_{thresh}. In reality, however, it is possible for events to translocate the system without ever filling the cavity. These events are referred to as straight-through events. The analysis in the previous sections is undefined for such events.

In practice, for the simulation parameters used here, the rate at which straight-through events occurs is always found to be well below 1%. As such, these events were simply discarded from the analysis. This is a source of bias, and straight-through events will be explored in more detail in future work.

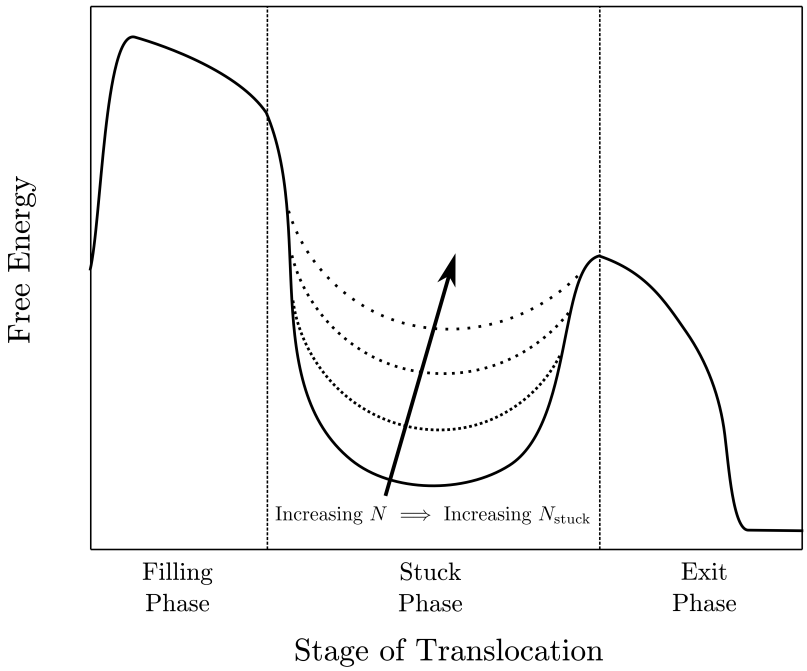

Figure 2.16: A schematic illustrating the qualitative features of the free energy landscape of CN translocation for $N < N^*$, where there is no tail. The dotted lines show the effect of increasing N. Since $N < N^*$, $N_{\text{stuck}} \approx N$, and increasing N increases N_{stuck}. This makes the free energy well of the stuck phase less deep, decreasing translocation time.

2.2.3 Free Energy Landscape for CN Translocation in the Basic Regime

In this section, the discussions of the previous sections will be summarized by describing the CN translocation process as the motion of the polymer through a free energy landscape. The landscape for translocation in the basic regime is shown in Figure 2.16 for $N < N^*$ and in Figure 2.17 for $N > N^*$. The free energy is presented as a one-dimensional function of the stage of translocation.

At each point in the landscape, the polymer experiences an effective force equal to the negative gradient of the free energy at that point. However, the

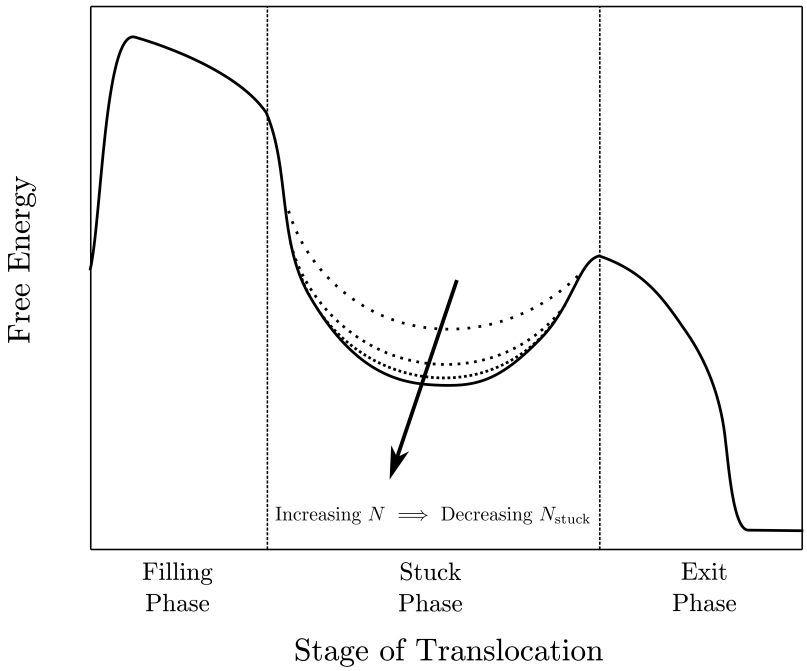

Figure 2.17: A schematic illustrating the qualitative features of the free energy landscape of CN translocation for $N > N^*$, where there is a tail. The dotted lines show the effect of increasing N. Since $N > N^*$, N_{stuck} decreases with N. This makes the free energy well of the stuck phase deeper, increasing translocation time.

particle is constantly subject to a random force corresponding to thermal motion. Thus, for instance, it is possible for the polymer to escape the stuck phase to enter the exit phase, despite the free energy barrier, if it experiences a sufficiently large random force in the right direction.

A particle moving in a free energy landscape will eventually access all parts of the landscape under the influence of thermal motion. The timescale on which a particle will move from a point a in the landscape to a point b is given by

$$\tau_{\text{landscape}} = \frac{1}{D}\int_a^b e^{A(y)/k_BT}dy \int_{-\infty}^{y} e^{A(x)/k_BT}dx, \tag{2.20}$$

where $A(x)$ is the free energy at point x in the landscape, T is the temperature of the particle, and D is the particle's diffusion coefficient [36, 72, 48, 9]. When considering the time for a particle to cross a large free energy barrier, like the one separating the stuck phase from the exit phase, this formula yields reduces to

$$\tau_{\text{escape}} \sim e^{\frac{A_b - A_a}{k_B T}}. \tag{2.21}$$

In other words, escape from a well in the free energy landscape grows exponentially with the depth of the well.

With this in mind, consider the features of CN translocation summarized in Figures 2.16 and 2.17. At the beginning of translocation, the polymer must cross a free energy barrier to begin the filling phase. Failure to cross this barrier corresponds to failed translocation events.

Once the polymer begins filling the cavity, it experiences a strong force pushing it to continue filling the cavity. It is rather unlikely for thermal fluctuations to reverse the filling of the cavity at this stage. The filling continues until the cavity is full and the polymer enters the stuck phase.

In the stuck phase, the polymer is trapped inside the cavity. It experiences large free energy barriers against exiting the cavity in either direction. The barrier is smaller towards the exit phase due to the action of the electric field. Since the escape time depends exponentially on the barrier height, a modest difference in the free energy barriers is sufficient to ensure the polymer will almost never retract back to *cis* from the stuck phase.

The depth of the well in the stuck phase depends on N_{stuck}. As established previously, adding monomers to the cavity increases the free energy cost of remaining in the cavity, so that the relative height of the free energy barrier to escape is reduced. This captures the decay of t_{stuck} with N when no tail exists

during the stuck phase.

When a tail forms, the barrier to escaping the stuck phase increases, manifested as a rapid increase in t_{stuck}. To a first approximation, this occurs because N_{stuck} decreases. This is shown in Figure 2.17. However, as discussed before and illustrated in Figure 2.10, the emergence of a tail increases t_{stuck} by a greater amount than can be attributed to the reduction of N_{stuck} alone. This is not illustrated in the landscape of Figure 2.17, as it is due to the change between open and closed confinement of the polymer in the cavity during the stuck phase, which cannot be easily represented schematically. What is shown in the landscape, however, is that, as N becomes very large, both N_{stuck} and t_{stuck} seem to converge to limiting value, so that the height of the free energy barrier must also converge to some limiting value.

After the exit phase, the polymer escapes into free solution and its free energy associated with the CN goes to zero.

Thus the free energy landscapes in Figures 2.16 and 2.17 capture most of the interesting properties of polymer translocation through the CN in the basic regime. The next sections will explore translocation for modified translocation parameters, specifically different cavity sizes, different applied force strengths, and different entrance/exit pore widths. The results of these analyses will be related back to the basic regime by showing modified free energy landscapes that express the changes in translocation dynamics.

2.3 Results of Varying Cavity Size in the Cavity-Nanopore

Figure 2.18 shows the translocation time as a function of chain length for two different CN cavity sizes. The blue line shows the data for the basic regime,

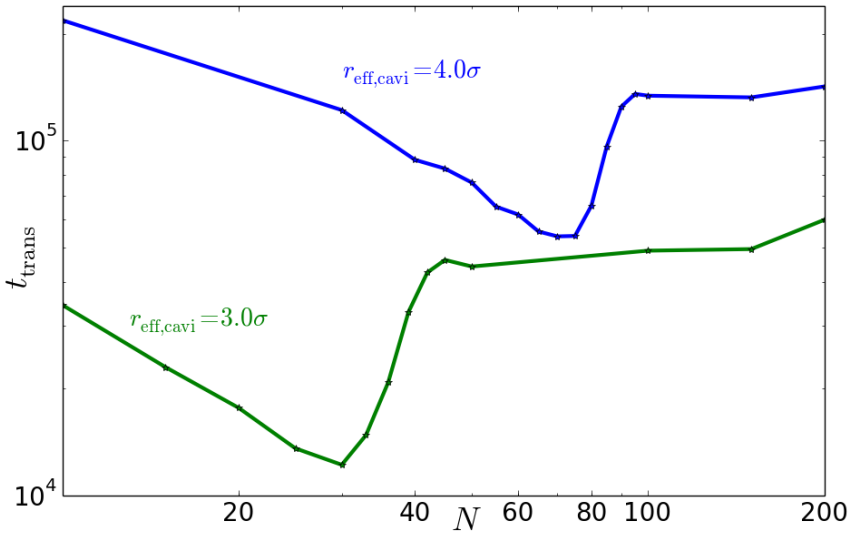

Figure 2.18: Plots of the translocation time, t_{trans}, for two different cavity sizes. In both cases, $F_{\text{pore}} = 0.4\epsilon/\sigma$ and $r_{\text{nom,pore}} = 1.3\sigma$. The blue line shows the basic regime, where $r_{\text{eff,cavi}} = 4.0\sigma$. The green line shows the data for a smaller cavity with $r_{\text{eff,cavi}} = 3.0\sigma$.

which is the same as the data in Figure 2.5. The green line shows the data for a CN with a small cavity, $r_{\text{eff,cavi}} = 3.0\sigma$. The applied force is kept the same in the entrance and exit pores, and the entrance/exit pore widths are kept the same as in the basic regime.

The data obtained for the smaller cavity size is consistent with the original result from the basic regime. The overall shape of the function is the same, with two major modifications. In the smaller cavity,

- The magnitude of t_{trans} is much smaller.
- The critical chain length N^* at which t_{trans} is minimal is much smaller.

These results are consistent with the discussion of translocation dynamics given in the previous sections.

When the cavity size is reduced, the entropic barrier to exit from the stuck

phase is reduced. This follows immediately, since a smaller cavity implies stronger polymer confinement, which increases the free energy cost of remaining in the stuck phase. As a result, t_{stuck} is reduced for the smaller cavity size. It is clear from the shape of the plot of t_{trans} in Figure 2.18 that the translocation time is still dominated by the stuck time for the $r_{\text{eff,cavi}} = 3.0\sigma$ case, so reducing t_{stuck} causes the reduction in t_{trans}.

It is also natural that the smaller cavity size lead to a smaller critical chain length. The discussion in the previous section indicates that the minimum of t_{trans} occurs shortly before the emergence of tails in the stuck phase. Tails emerge in the stuck phase when N is too large for the entire polymer to fit inside the cavity during the filling phase. Clearly the largest polymer that can fit inside a cavity without forming a tail will be longer for larger cavities. In other words, N^* is expected to be an increasing function of $r_{\text{eff,cavi}}$, which is indeed what is seen in Figure 2.18.

Beyond these two adjustments, the translocation time as a function of chain length does not change significantly when the $r_{\text{eff,cavi}}$ is changed from 4.0σ to 3.0σ. Simulations were also conducted at 3.5σ (not shown), and the conclusion is the same. This supports the claim that the description of CN translocation dynamics given in the basic regime continues to hold for a range of cavity sizes.

In terms of the free energy landscapes shown in Figures 2.16 and 2.17, changing the cavity size by a small amount simply rescales the dimensions of the landscape. Smaller cavities correspond to shallow wells in the stuck phase. Furthermore, the total duration of the translocation process is reduced. However, the shape of the landscape remains effectively the same as for the basic regime.

2.4 Results of Increasing Voltage in the Cavity-Nanopore

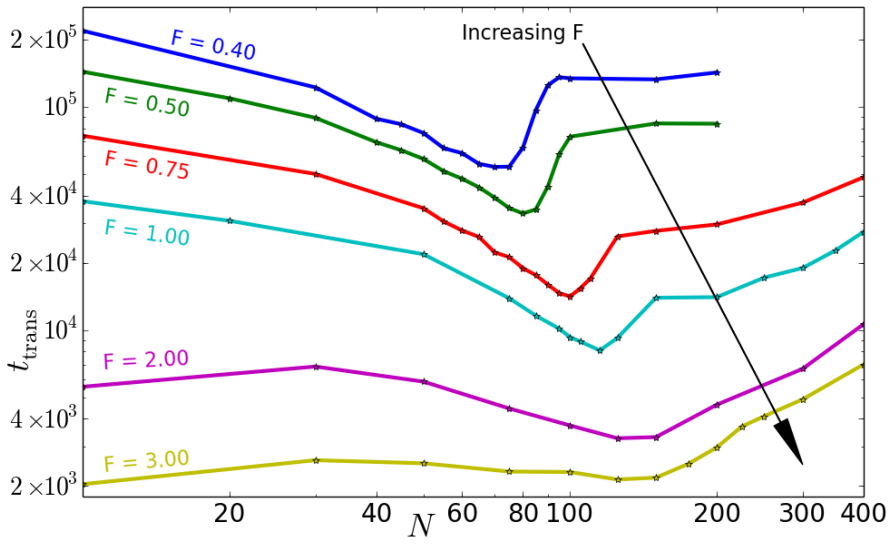

Figure 2.19: Plots of the translocation time, t_{trans}, for a range of applied forces. In both cases, $r_{\mathrm{eff,cavi}} = 4.0\sigma$ and $r_{\mathrm{nom,pore}} = 1.3\sigma$. The topmost line (blue) corresponds to the basic regime, as in Figure 2.5. The other lines are labelled according to F_{pore}, and the arrow indicates the direction of increasing force.

Figure 2.19 shows the translocation time as a function of chain length for a range of applied forces. The blue line shows the data for the basic regime, which is the same as the data in Figure 2.5. The cavity size and entrance/exit pore widths are kept constant.

The most evident result is that the translocation time decreases rapidly as a function of applied field strength. For instance, increasing the force by a factor of two from $F = 0.5\epsilon/\sigma$ to $F = 1.0\epsilon/\sigma$ decreases t_{trans} at $N = 10$ by almost a factor of four. Increasing by another factor of two from $F = 1.0\epsilon/\sigma$ to $F = 2.0\epsilon/\sigma$ decreases t_{trans} at $N = 10$ by over a factor of six. Whereas translocation time generally scales linearly with applied force in the SN system,

the translocation time in the CN system decreases much more rapidly with increasing force. Furthermore, the translocation time decreases more rapidly with increasing force at higher forces.

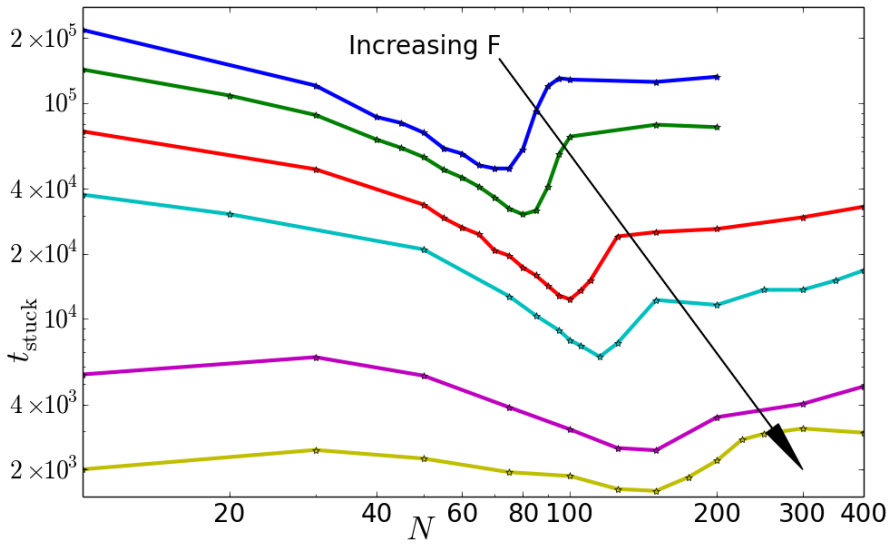

Figure 2.20: Plots of the stuck time, t_{stuck}, for a range of applied forces. In both cases, $r_{\text{eff,cavi}} = 4.0\sigma$ and $r_{\text{nom,pore}} = 1.3\sigma$. The topmost line (blue) corresponds to the basic regime, as in Figure 2.7. The other lines are labelled according to F_{pore}, and the arrow indicates the direction of increasing force.

Since the filling and exit phases somewhat resemble translocation through SN systems, it is natural to suspect that this non-linear dependence of t_{trans} on the applied force is related to the stuck phase. Figure 2.20 shows t_{stuck} for the same range of applied forces as shown in Figure 2.19. As expected, most of the features seen in t_{trans} are present in t_{stuck}. Thus the non-linear scaling of t_{trans} with F_{pore} is indeed due to the effect of the force on the stuck phase.

At low forces, the t_{stuck} plot does not change appreciably with increasing force. The critical chain length N^* grows with increasing force. Recall that the previous discussion implied that N^* corresponds to the onset of tails during the stuck phase. Tails form when the eletric force pushing monomers into the

cavity cannot overcome the entropic force associated with the free energy cost of confining a polymer. As the applied force is increased, a larger entropic force can be overcome, and so larger chains can be forced entirely into the cavity during the stuck phase. Thus tails emerge at larger values of N, which explains why N^* increases with greater applied force.

At higher forces, neither t_{trans} nor t_{stuck} displays the pronounced minimum that was characteristic of the system's behaviour in the basic regime. In fact, for $F_{\text{pore}} = 3.0\epsilon/\sigma$, the plot of t_{stuck} is virtually constant for all chains lengths below some threshold, in this case $N \approx 200$. It is not obvious from Figures 2.19 and 2.20 alone whether this threshold is the high-force analogue of the critical chain length.

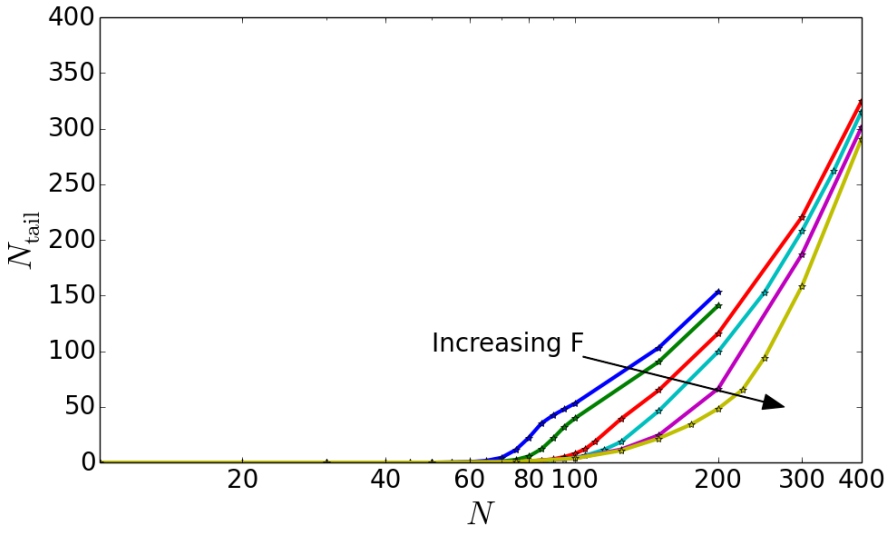

Figure 2.21: Plots of the average tail size, N_{tail}, for a range of applied forces. In both cases, $r_{\text{eff,cavi}} = 4.0\sigma$ and $r_{\text{nom,pore}} = 1.3\sigma$. The topmost line (blue) corresponds to the basic regime, as in Figure 2.7. The other lines are labelled according to F_{pore}, and the arrow indicates the direction of increasing force.

Figure 2.21 shows the average tail size during the stuck phase, N_{tail}, as a function of N over a range of applied forces. For the very high forces, it is

not very clear that the the emergence of the tail is directly responsible for the threshold effect in t_{stuck}, since N_{tail} is non-zero for chains much shorter than the threshold lengths observed in the t_{stuck} plots. One can argue as follows, however: the sharp rise in t_{trans} in the basic regime ($F_{\text{pore}} = 0.4\epsilon/\sigma$) corresponds to $N_{\text{tail}} \approx 50$. When $F_{\text{pore}} = 3.0\epsilon/\sigma$, the tail achieves this length of $N_{\text{tail}} \approx 50$ at $N \approx 200$. In turn, the threshold in t_{stuck} also occurs at $N \approx 200$. In other words, this argument suggests that the emergence of the tail is still responsible for the threshold effect in the high-force limit, and simply that $N_{\text{tail}} > 0$ is not a perfect indication of where the tail actually starts to contribute appreciably to the system dynamics.

In summary, increasing the applied force decreases the stuck time rapidly. For moderate increases in the force, the qualitative behaviour of the system remains the same as in the basic regime, but N^* increases with force. At very large forces, the qualitative behaviour of the system changes: the stuck time is constant below some threshold chain length, then increases to a new plateau beyond the threshold. Figure 2.21 can be interpreted to conclude that the emergence of a tail in the stuck phase is responsible for both the minimum in t_{stuck} at low forces and the threshold effect in t_{stuck} at higher forces.

If the emergence of the tail is responsible for the threshold effect at higher forces, then chains shorter than this threshold correspond to those than can fit entirely inside the cavity, on average. Thus, the fact that t_{stuck} is constant for chains shorter than the threshold implies that t_{stuck} is constant for chains that can fit entirely inside the cavity. This effect can be understood by realizing that at large values of F_{pore}, the force inside the cavity F_{cavi} is no longer negligible.

This force field confines the polymer towards the *trans* wall of the cavity. This significantly reduces the free energy barrier to escape, as the force itself encourages threading through the exit pore, and now even short polymers are

tightly confined against the *trans* wall. The field acts on each monomer of the polymer independently, so the net force on the polymer is a function of N. Thus the free energy barrier to escape is reduced by a different amount for different chain lengths. This length-dependence can be balanced against the other length-dependence of the barrier, so that the free energy barrier to escape can be made independent of N. This explains how t_{stuck} is roughly constant with N below the threshold when $F_{\text{pore}} = 3.0\epsilon/\sigma$.

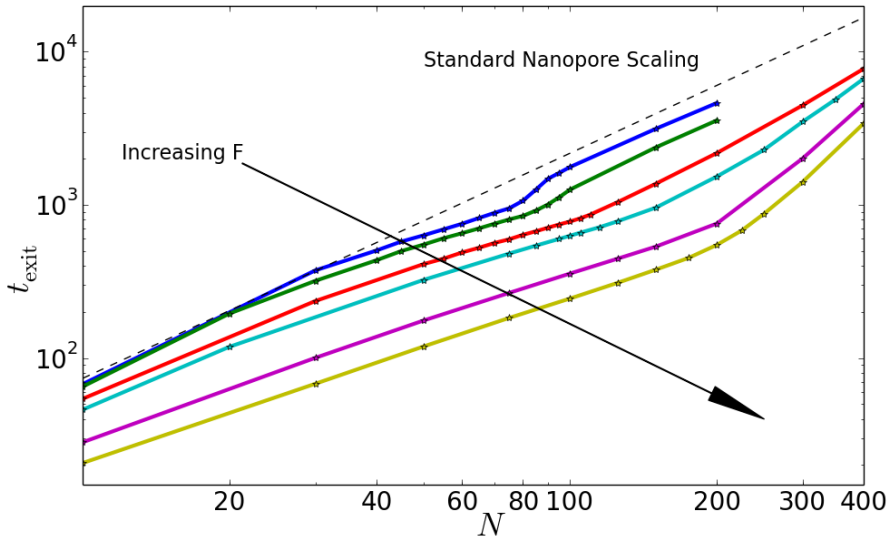

Figure 2.22: Plots of the exit time, t_{exit}, for a range of applied forces. In both cases, $r_{\text{eff,cavi}} = 4.0\sigma$ and $r_{\text{nom,pore}} = 1.3\sigma$. The topmost line (blue) corresponds to the basic regime, as in Figure 2.11. The other lines are labelled according to F_{pore}, and the arrow indicates the direction of increasing force.

At large N and at higher forces, t_{trans} no longer agrees with t_{stuck}. The difference is predominantly due to the exit time, shown in Figure 2.22. As F_{pore} increases, t_{stuck} decreases much faster than t_{exit}, so the exit time becomes a larger fraction of the total translocation time. Since the exit time is always monotonically increasing with N, this causes t_{trans} to increase with N at large N and higher F_{pore}.

The effects of the electric field on the free energy landscape are summarized in Figure 2.23. As the field strength increases, the landscape is tilted by a greater amount. This tilting reduces the barrier to escape from the stuck phase, which causes a rapid decrease in t_{stuck}, since this time depends exponentially on the barrier height.

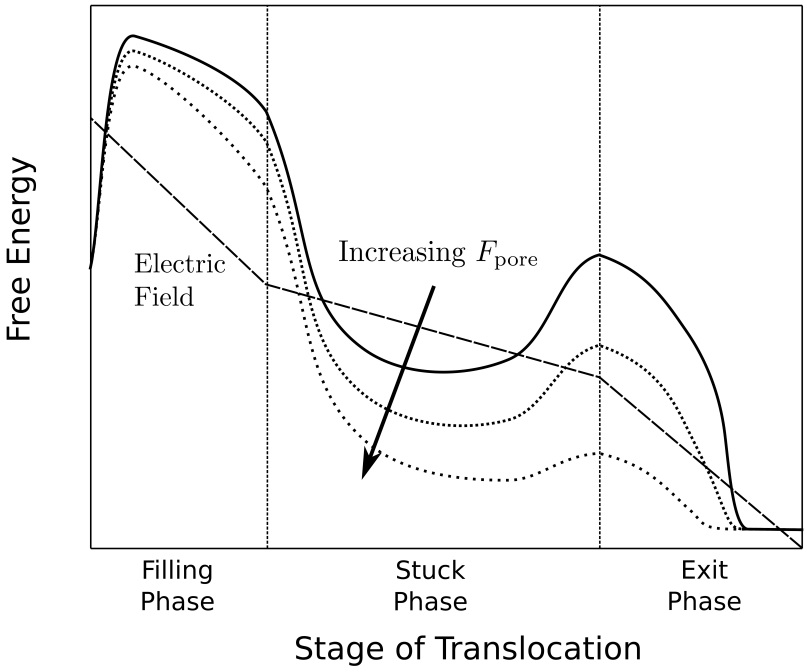

Figure 2.23: A schematic illustrating the qualitative effects of varying F_{pore} on the shape of the free energy landscape. The arrow indicates the direction of increasing force, and the dashed line indicates the general shape of the electric potential energy. Increasing the force tilts the landscape, lessening the barrier to escape, and thus rapidly decreasing t_{trans}.

2.5 Results of Wider Pores in the Cavity-Nanopore

This section explores the effect of changing the entrance/exit pore width while keeping the cavity size and applied force strength constant. However, data

was collected at $F_{\text{pore}} = 0.5\epsilon/\sigma$, rather than $F_{\text{pore}} = 0.4\epsilon/\sigma$ as in the basic regime. As the previous section demonstrates, changing the force by this amount does not alter the simulation results significantly. The discussion of system dynamics for the basic regime applies equally well to the case with $r_{\text{eff,cavi}} = 4.0\sigma, r_{\text{nom,pore}} = 1.3\sigma$, and $F_{\text{pore}} = 0.5\epsilon/\sigma$, which is used as the reference case in this section.

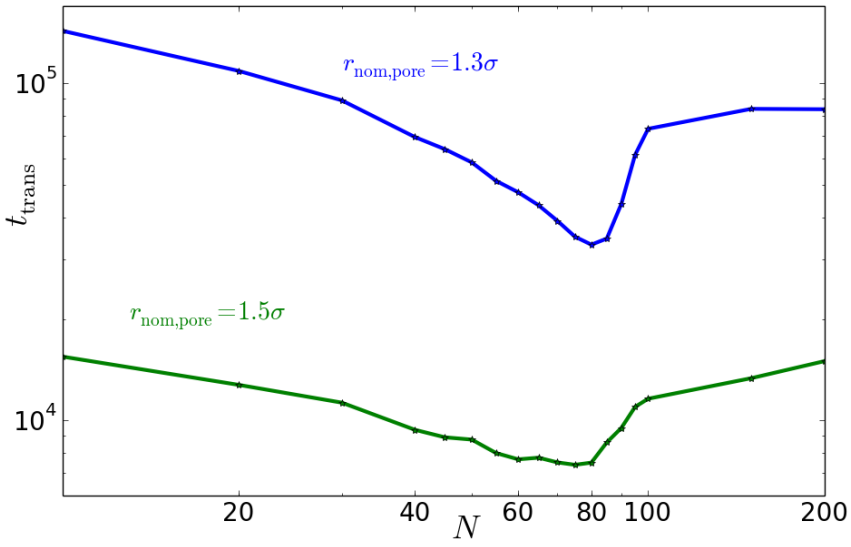

Figure 2.24: Plots of the translocation time, t_{trans}, for two different entrance/exit pore widths. In both cases, $F_{\text{pore}} = 0.5\epsilon/\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. The blue line shows the results for the reference case, where $r_{\text{nom,pore}} = 1.3\sigma$. The green line shows the data obtained for the wider pore, where $r_{\text{nom,pore}} = 1.5\sigma$.

Figure 2.24 shows the translocation time as a function of chain length for two different CN entrance/exit pore widths. The blue line shows the data for the reference case, which is quite similar to the data in Figure 2.5. The green line shows the data for a CN with wider entrance/exit pores, $r_{\text{nom,pore}} = 1.5\sigma$.

The plot for the case with wider pores differs significantly from the data in the reference case at larger N values, although the two cases are very similar for smaller N values. For short chains, the translocation time in the wider pore

case is smaller by an order of magnitude, but otherwise scales in roughly the same way as the translocation time in the basic regime.

For larger N values, these discrepancies are evident:

- The minimum of the translocation time occurs at a shorter chain length, i.e. larger $r_{\text{nom,pore}}$ leads to smaller N^*.
- The well in the translocation time around N^* is not as deep in the case with the wider pore.
- The sharp increase in t_{trans} as N increases beyond N^* is not as sharp in the case with the wider pore.
- After $N > N^*$, whereas t_{trans} plateaus in the reference case, in the case with the wider pore t_{trans} continues to grow at an appreciable rate.

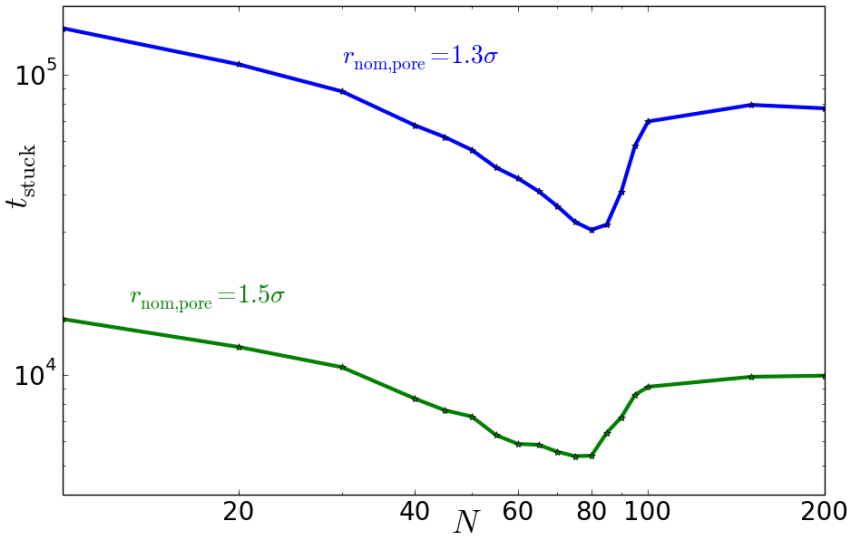

Figure 2.25: Plots of the stuck time, t_{stuck}, for two different entrance/exit pore widths. In both cases, $F_{\text{pore}} = 0.5\epsilon/\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. The blue line shows the results for the reference case, where $r_{\text{nom,pore}} = 1.3\sigma$. The green line shows the data obtained for the wider pore, where $r_{\text{nom,pore}} = 1.5\sigma$.

To understand these discrepancies, first consider t_{stuck}, as we have so far assumed that $t_{\mathrm{trans}} \approx t_{\mathrm{stuck}}$. Figure 2.25 compares t_{stuck} between the basic regime and the case with wider pores. This plot demonstrates that the shifted, wider, and shallower well around the minimum value of t_{trans} occurs in t_{stuck} as well, so that these deformations are probably attributable to dynamics in the stuck phase. However, t_{stuck} does not grow at large N values, so this discrepancy is due to dynamics outside the stuck phase.

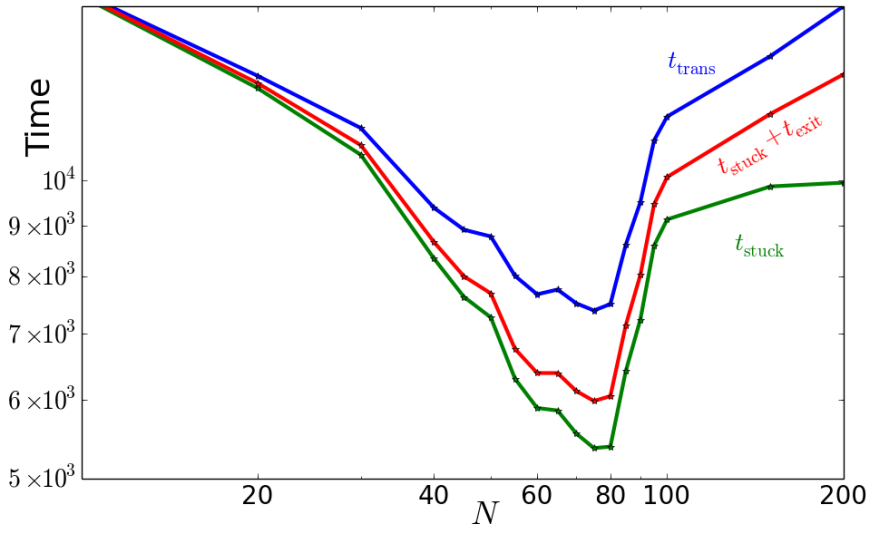

Figure 2.26: Plots of t_{trans} (topmost line, blue), t_{stuck} (bottommost line, green), and $t_{\mathrm{stuck}} + t_{\mathrm{exit}}$ (middle line, red) for the case with the wider pore ($F_{\mathrm{pore}} = 0.5\epsilon/\sigma, r_{\mathrm{eff,cavi}} = 4.0\sigma, r_{\mathrm{nom,pore}} = 1.5\sigma$).

Since, in the case with the wider pore, t_{trans} continue to increase at large N but t_{stuck} does not, the increase must be due to t_{fill} and t_{exit}. Figure 2.26 compares t_{trans}, t_{stuck}, and $t_{\mathrm{stuck}} + t_{\mathrm{exit}}$ for the case with the wider pore. The difference between t_{trans} and $t_{\mathrm{stuck}} + t_{\mathrm{exit}}$ is t_{fill}, by definition.

At large N values, the plot of t_{trans} is parallel to the plot of $t_{\mathrm{stuck}} + t_{\mathrm{exit}}$. Thus t_{exit} is responsible for the growth of t_{trans} at large N values. Specifically, t_{trans} grows at large N values in the case with the wider pore because t_{exit} is

no longer a negligible component of t_{trans} under these conditions. This occurs because the widening of the pore decreases t_{stuck} much more rapidly than it decreases t_{exit}.

In particular, Figure 2.26 illustrates that neither t_{fill} nor t_{exit} can be neglected in this wider pore case. The filling phase has not been studied in detail at this point, so further analysis is required to understand its contribution to the dynamics in the wide pore case.

Nonetheless, the plots of t_{stuck} in Figure 2.25 do imply that the shifted, wider, and shallower minimum of t_{trans} in the wider pore case are attributable to the stuck phase. Recall that the minimum arises because chains much shorter than N^* do not form tails, whereas chains much larger than N^* do form tails. This suggests that, in the case with the wider pore, tails form at a lower value of N, so N^* occurs at a lower value.

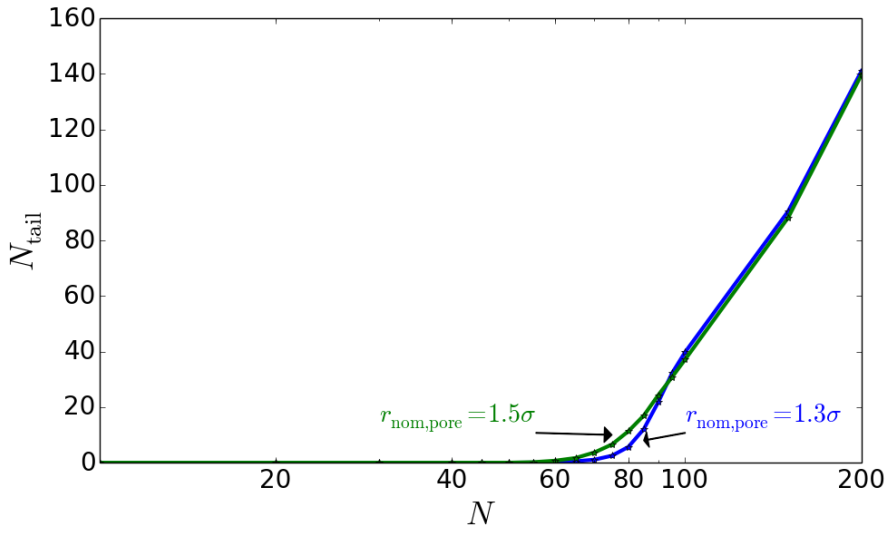

Figure 2.27: Plots of the average tail size during the stuck phase, N_{tail}, for two different entrance/exit pore widths. In both cases, $F_{\text{pore}} = 0.4\epsilon/\sigma$ and $r_{\text{eff,cavi}} = 4.0\sigma$. The blue line shows the results for the reference case, where $r_{\text{nom,pore}} = 1.3\sigma$. The green line shows the data obtained for the wider pore, where $r_{\text{nom,pore}} = 1.5\sigma$.

Figure 2.27 compares the plots of N_{tail} between the reference case and the case with the wider pore. This confirms that tails emerge at shorter N values in the case with the wider pore, which accounts for the shifting of N^* to lower values. The minimum in t_{stuck} is shallower and wider because, since tails emerge sooner, the difference between the stuck states with and without tails is not as pronounced.

This accounts for all of the features observed in Figure 2.24. In summary, widening the entrance/exit pores affects the stuck phase significantly. It reduces the stuck time significantly, so that at large N the contribution of t_{exit} is sufficient to make t_{trans} grow with N. Near the critical chain length, the wider pore facilitates the emergence of tails during the stuck phase. This means N^* occurs at a smaller value, and the well in t_{trans} around N^* is not as deep nor as sharp in the case with the wider pores.

2.6 A Free Energy Model to Predict the Critical Chain Length

The detailed insight into the translocation dynamics afforded by the analysis in the previous sections suggests a relatively simple model for predicting the critical chain length N^* at which t_{trans} experiences such a sharp minimum. This minimum is most pronounced in the narrow pore, low-F_{pore} limit. The following results have been established for that regime:

1. As per Figures 2.5 and 2.7, $t_{\text{trans}} \approx t_{\text{stuck}}$.
2. As per Figure 2.10, the N that minimizes t_{stuck} also maximizes N_{stuck}.
3. As per Figures 2.8 and 2.9, the maximum of N_{stuck} occurs before tails are very important.

These will be referred to as the Basic Results.

Using these results, a simple model will be constructed to estimate the critical chain length N^*. The model will predict the chain length N that maximizes N_{stuck}, as this corresponds to N^* according to Basic Results 1 and 2. Since N_{stuck} is defined in the stuck phase, the model will only consider polymers in the stuck state. Finally, the model will assume that no tail is present in the stuck state, as per Basic Result 3.

The toy model will consist simply of a polymer at equilibrium confined entirely inside the cylindrical CN cavity. The entrance pore is ignored since the polymer has no tail, and the exit pore is ignored since the polymer is in the stuck state. However, the polymer is considered to have entered the cavity via the entrance pore, and gained energy from the electric field in so doing. This energy is balanced against the free energy cost of confining the polymer inside the cavity. Starting with a short chain, we will increase the length of the polymer one monomer at a time until the free energy cost of increasing the length of the confined polymer exceeds the free energy lost by adding monomers to the cavity via the entrance pore. The chain length at which this occurs will be the estimate for the chain length that maximizes N_{stuck}.

Calculating this chain length is now just a matter of obtaining and minimizing an expression for the total free energy of the system as a function of chain length. Such an expression, however, is non-trivial. An approximate free energy expression, obtained using some simplifying assumptions, is now derived.

Adding monomers to the cavity decreases the free energy due to the electric field. Figure 2.28 illustrates a monomer losing free energy as it enters the cavity. Each monomer gains an energy $A_{\text{pore}} = F_{\text{pore}}\sigma$ crossing the entrance pore, and then gains some energy from the field inside the cavity based on its exact position. Whereas the first energetic term, due to crossing the entrance

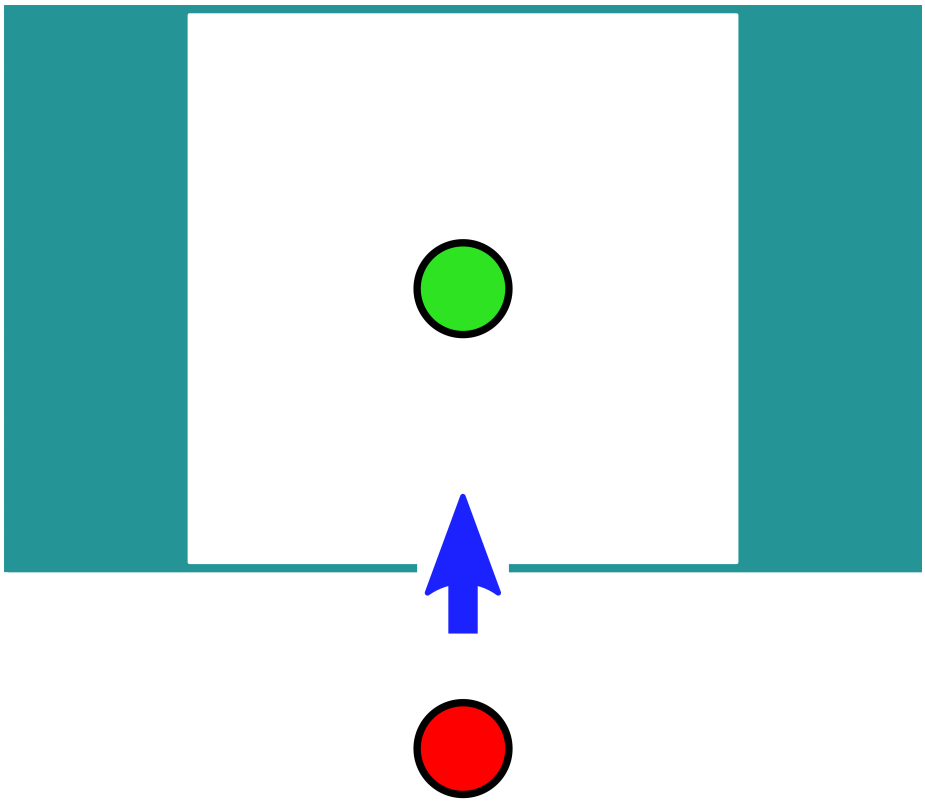

Figure 2.28: Schematic illustrating a monomer losing free energy to the electric field as it crosses the entrance pore. The red monomer has more free energy than the green monomer.

pore, is the same for each monomer in the cavity, the second term requires detailed knowledge of the conformation of the polymer inside the cavity. This information is difficult to obtain.

Call the energy of a monomer due to its position in the electric field inside the cavity A_{cavi}. This term is maximized when the monomer is touching the *trans* wall of the cavity, so

$$\max\left(A_{\text{cavi}}\right) = F_{\text{cavi}} h_{\text{eff,cavi}}, \tag{2.22}$$

where $h_{\text{eff,cavi}}$ is the effective height of the cavity. The force in the cavity satisfies

$$F_{\text{cavi}} = F_{\text{pore}} \left(\frac{r_{\text{pore}}}{r_{\text{cavi}}}\right)^2, \tag{2.23}$$

so the bound on A_{cavi} can be rewritten as

$$\max\left(A_{\text{cavi}}\right) = F_{\text{pore}} h_{\text{eff,cavi}} \left(\frac{r_{\text{eff,pore}}}{r_{\text{eff,cavi}}}\right)^2. \tag{2.24}$$

The cavity is specified to have a height equal to its diameter, i.e. $h_{\text{eff,cavi}} = 2r_{\text{eff,cavi}}$. So the bound can be rewritten again as

$$\max\left(A_{\text{cavi}}\right) = F_{\text{pore}}(2r_{\text{eff,cavi}}) \left(\frac{r_{\text{eff,pore}}}{r_{\text{eff,cavi}}}\right)^2 = 2F_{\text{pore}} r_{\text{eff,pore}} \left(\frac{r_{\text{eff,pore}}}{r_{\text{eff,cavi}}}\right). \tag{2.25}$$

In the basic regime, $r_{\text{eff,cavi}} = 4.0\sigma$ and $r_{\text{eff,pore}} = 0.8\sigma$. In that case, the bound becomes

$$\max\left(A_{\text{cavi}}\right) = 0.32 F_{\text{pore}} \sigma = 0.32 A_{\text{pore}}. \tag{2.26}$$

Thus the electric field inside the cavity is much less important than the electric field in the entrance pore. However, it is still possible that the electric field in the cavity could significantly affect the conformation of the polymer in the stuck state. To test this, the energy acquired from the electric field must be compared to the average thermal energy. For a single monomer,

$$\max\left(A_{\text{cavi}}\right) = 0.32 A_{\text{pore}} = 0.32 F_{\text{pore}} \sigma. \tag{2.27}$$

The field in the basic regime was $F_{\text{pore}} = 0.4\epsilon/\sigma$. Since the simulations used $\epsilon = k_B T$,

$$\max\left(E_{\text{cavi}}\right) = 0.32(0.4\epsilon) = 0.128 k_B T. \tag{2.28}$$

Thus the maximum energy of interaction between each monomer and the electric field in the cavity is an order of magnitude weaker than the average monomer thermal energy, and thus this field is unlikely to significantly impact the polymer's conformation in this low force regime.

Furthermore, this bound is the maximum possible value for A_{cavi}, which is only achieved when the monomer is against the *trans* wall of the cavity. Since the electric field in the cavity is weak compared to the thermal motion of the chain, the COM of the polymer is likely in the middle of the cavity on average, so the average value of E_{cavi} will be half the value of this upper bound:

$$\max\left(A_{\text{cavi}}\right) \ll k_B T \implies \langle A_{\text{cavi}} \rangle \approx F_{\text{cavi}} \frac{h_{\text{eff,cavi}}}{2} = \frac{1}{2} \max\left(A_{\text{cavi}}\right). \tag{2.29}$$

This relationship can then be inverted to estimate when the energy of interaction with the electric field in the cavity will become comparable to the thermal energy

of the polymer. The two are comparable when

$$\langle A_{\text{cavi}} \rangle \geq k_B T, \tag{2.30}$$

$$\frac{1}{2} 0.32 F_{\text{pore}} \sigma \geq k_B T, \tag{2.31}$$

$$F_{\text{pore}} \geq 6.25 \frac{\epsilon}{\sigma}. \tag{2.32}$$

In other words, in the low force limit, the energy due to monomer positions within the cavity is roughly an order of magnitude smaller than the energy due to monomers crossing the entrance pore. As such, the toy model will neglect the electric field inside the cavity. The total free energy lost due to the electric field is therefore

$$A_{\text{electric}} \approx -N F_{\text{pore}} \sigma. \tag{2.33}$$

This approximation is valid at $F_{\text{pore}} = 0.4\epsilon/\sigma$, and becomes insufficient at some point before $F_{\text{pore}} \approx 8\epsilon/\sigma$.

Conversely, adding monomers to the cavity increases the total free energy via the entropic cost of confining the polymer. This is illustrated in Figure 2.29. This free energy of confinement is difficult to compute. Cacciuto and Luijten have conducted a thorough study of precisely this free energy for a polymer in spherical confinement in the absence of forces. In the low-density regime, they find

$$A_{\text{entropic}} = B \left(\frac{R_G(N)}{r_{\text{eff,cavi}}} \right)^{\frac{3}{3\nu - 1}}, \tag{2.34}$$

for some constant B. A value of $B \approx 5$ was extracted graphically from their manuscript, using the online plot digitizer WebPlotDigitizer [58].

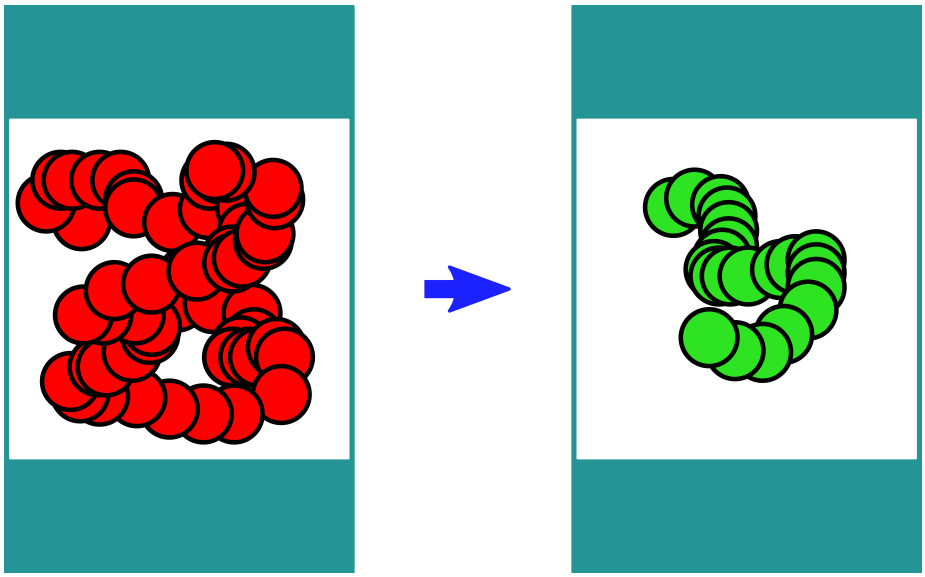

Figure 2.29: Schematic illustrating the entropic cost of confining a large polymer. The red polymer has more free energy than the green polymer.

Thus, the total free energy of the model system is

$$A_{\text{total}} = A_{\text{electric}} + A_{\text{entropic}} \tag{2.35}$$

$$\approx -NF_{\text{pore}}\sigma + B\left(\frac{R_G(N)}{r_{\text{eff,cavi}}}\right)^{\frac{3}{3\nu-1}}. \tag{2.36}$$

For the freely-jointed chain with excluded volume, the radius of gyration is of the form $R_G = C_G N^\nu \sigma$ for some scaling coefficients C_G, ν. So the free energy can be rewritten in terms of only N as

$$A_{\text{total}}(N) \approx -NF_{\text{pore}}\sigma + B\left(\frac{C_G N^\nu \sigma}{r_{\text{eff,cavi}}}\right)^{\frac{3}{3\nu-1}}, \tag{2.37}$$

which can be minimized by setting its derivative to zero:

$$\frac{dA_{\text{total}}}{dN} = 0 \tag{2.38}$$

$$\iff \qquad F_{\text{pore}}\sigma = B\left(\frac{C_G\sigma}{r_{\text{eff,cavi}}}\right)^{\frac{3}{3\nu-1}} \frac{d\left(N^{\frac{3\nu}{3\nu-1}}\right)}{dN} \tag{2.39}$$

$$\iff \qquad \frac{3\nu}{3\nu-1}N^{\frac{1}{3\nu-1}} = \frac{F_{\text{pore}}\sigma}{B}\left(\frac{r_{\text{eff,cavi}}}{C_G\sigma}\right)^{\frac{3}{3\nu-1}} \tag{2.40}$$

$$\iff \qquad N^*_{\text{theo}} = \left[\frac{3\nu-1}{3\nu}\frac{F_{\text{pore}}\sigma}{B}\left(\frac{r_{\text{eff,cavi}}}{C_G\sigma}\right)^{\frac{3}{3\nu-1}}\right]^{3\nu-1}. \tag{2.41}$$

The values of ν, C_G, and B used in Equation 2.41 must be measured from the simulations. However, the parameters ν and C_G are only measured once, as they depend on the numerical implementation of the polymer, not on any aspects of the CN system. The values used here were $\nu \approx 0.628, C_G \approx 0.402$, as given in the introduction.

The parameter B must be fit to the critical chain lengths measured from simulation. A value of $B \approx 4$ was found to fit the data most effectively. This fitted value of B is in good agreement with the value of $B \approx 5$ extracted from the paper by Cacciuto and Luijten, lending credence to Equation 2.41. The discrepancy arises in part because that paper studied spherical confinement, whereas the CN cavity is cylindrical.

The toy model, and therefore Equation 2.41, were derived in the low field limit. As shown above, at sufficiently high forces the conformation of the polymer inside the cavity will become affected by the electric field inside the cavity, which is currently neglected. However, this is expected to occur around $F_{\text{pore}} = 8\epsilon/\sigma$. Conversely, Cacciuto and Luijten found that the free energy of confinement changes to a different form when the polymer density inside the cavity exceeds a volume fraction of $\phi \approx 0.15$. In the present simulations, this occurs at a lower force than $F_{\text{pore}} = 8\epsilon/\sigma$, and so $\phi = 0.15$ will be considered

an effective upper bound for the applicability of the model.

Conversely, the term describing the free energy of confinement is also only applicable for confined polymers. Chains that are much smaller than the cavity cannot be considered confined. Specifically, for under-confined chains the free energy of confinement will be smaller than the relation found by Cacciuto and Luijten. As a result, Equation 2.41 will overpredict the cost of confinement, and thus underpredict the critical chain length. Chains will become under-confined when $R_G \approx r_{\text{eff,cavi}}$. This will be considered an effective lower bound for the applicability of the model.

Figure 2.30 shows the comparison between Equation 2.41 and the values of N^* that were found to minimize t_{trans} in simulations. The top plot is for the geometry with the narrower pore, $r_{\text{nom,pore}} = 1.3\sigma$, and the bottom plot is for the case with the wider pore, $r_{\text{nom,pore}} = 1.5\sigma$. The minima were collected for cavity sizes of $r_{\text{eff,cavi}} = 3.0\sigma, 3.5\sigma$, and 4.0σ, and for a range of field strengths. The theoretical upper and lower bounds are also shown on the plots.

The agreement is good in the narrow pore, low-F_{pore} cases between the bounds of applicability. The point at $r_{\text{eff,cavi}} = 3.0\sigma$ and $F_{\text{pore}} = 0.4\epsilon/\sigma$ lies along the lower bound of applicability, supporting the credibility of this bound. The prediction for $F_{\text{pore}} = 0.6\epsilon/\sigma$ and $0.75\epsilon/\sigma$ fit the data rather poorly, despite being mostly within the bounds of applicability. This will be addressed later.

Conversely, the wide pore case in Figure 2.30 does not agree with the prediction at all. This disagreement arises because Equation 2.41 was derived from Basic Result 1, namely that $t_{\text{trans}} \approx t_{\text{stuck}}$. The model completely neglects any contributions from the filling and exit phases. In the wide pore case, this is no longer the case, as per Figure 2.26 which shows the contributions of each phase to t_{trans}.

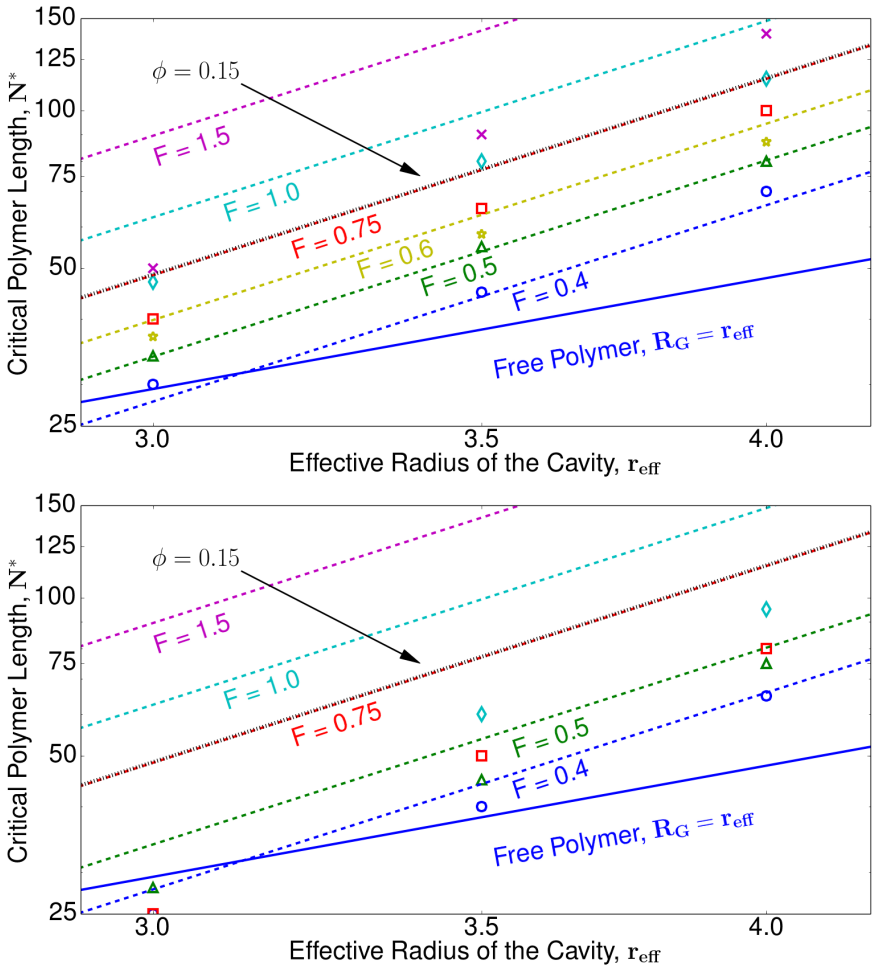

Figure 2.30: Comparison of the free energy model for $r_{\mathrm{nom,pore}} = 1.3\sigma$ (top) and $r_{\mathrm{nom,pore}} = 1.5\sigma$ (bottom) to the values of N measured from simulation to minimize t_{trans}.

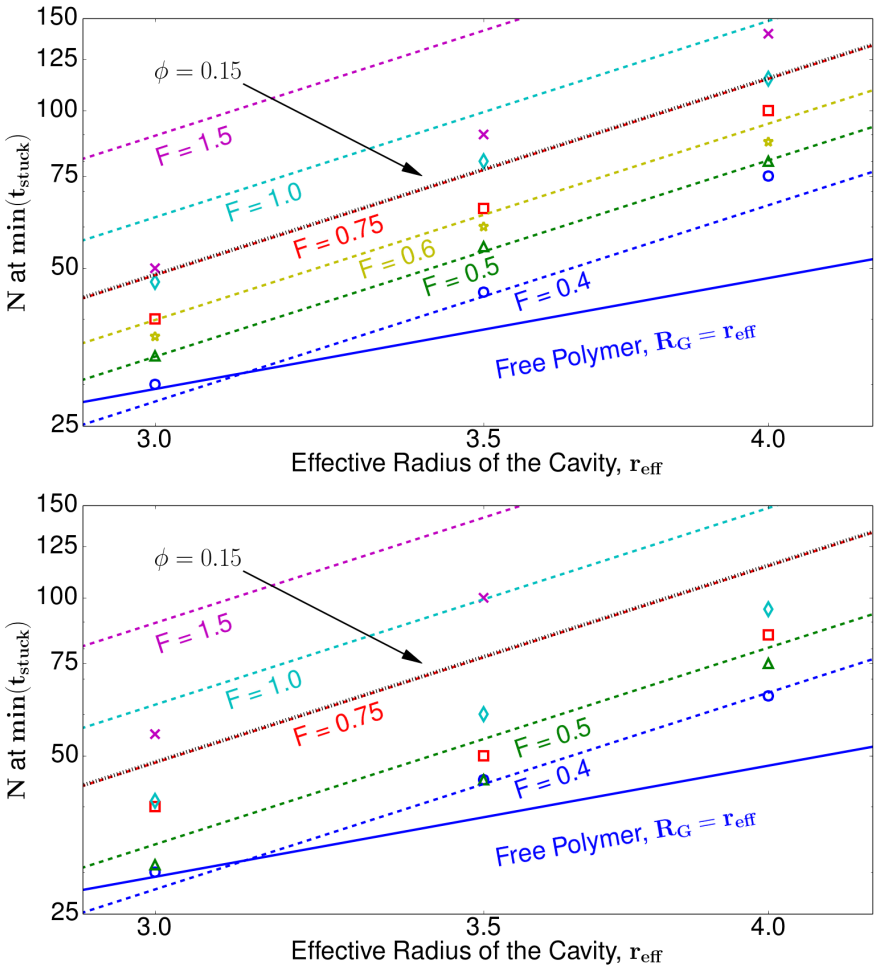

Figure 2.31: Comparison of the free energy model for $r_{\mathrm{nom,pore}} = 1.3\sigma$ (top) and $r_{\mathrm{nom,pore}} = 1.5\sigma$ (bottom) to the values of N measured from simulation to minimize t_{stuck}.

To test the magnitude of this effect, Figure 2.31 shows the comparison between the predictions of Equation 2.41 and the values of N measured from simulation to minimize t_{stuck}. The agreement is much better for the wider pore case, especially at $F_{\text{pore}} = 0.4\epsilon/\sigma$. However, the agreement in the wider pore case is still rather poor for higher forces.

The data from the case with the wide pore is still in poor agreement with the prediction because it violates Basic Result 2, namely that the minimum of t_{stuck} coincides with the maximum of N_{stuck}. Figure 2.32 shows plots of t_{stuck} against N_{stuck} for the narrow and wide pore cases at $F_{\text{pore}} = 0.4\epsilon/\sigma, 0.5\epsilon/\sigma$, and $0.75\epsilon/\sigma$. In the narrow pore case, the minima of t_{stuck} and the maxima of N_{stuck} coincide perfectly for the two lowest forces. On the plot, this corresponds to the sharpness of the line where it reverses direction. Even the third, highest force case has a relatively sharp plot, i.e. the maximum of t_{stuck} and the maximum of N_{stuck} are still close to one another.

This is not true for the case with the wider pore: the end of the plots are not sharp even at the lowest force. In other words, the minima of t_{stuck} and the maxima of N_{stuck} are much less close at the same force when the entrance/exit pores are widened. Thus Basic Result 2 does not hold, and Equation 2.41 does not fit the data well.

To confirm that this is responsible for the discrepancy in the fit of the wide pore data, Figure 2.33 shows the comparison between Equation 2.41 and the values of N that maximize N_{stuck}. The quality of the fit in the wide pore case is now comparable to that in the narrow pore case. Furthermore, the fit has improved in the narrow pore case at higher forces. In particular, the cases at $F_{\text{pore}} = 0.6\epsilon/\sigma, 0.75\epsilon/\sigma$ are now in much better agreement.

In other words, Equation 2.41 accurately predicts the length N that maximizes N_{stuck}. This chain length only corresponds to the true N^* when consider-

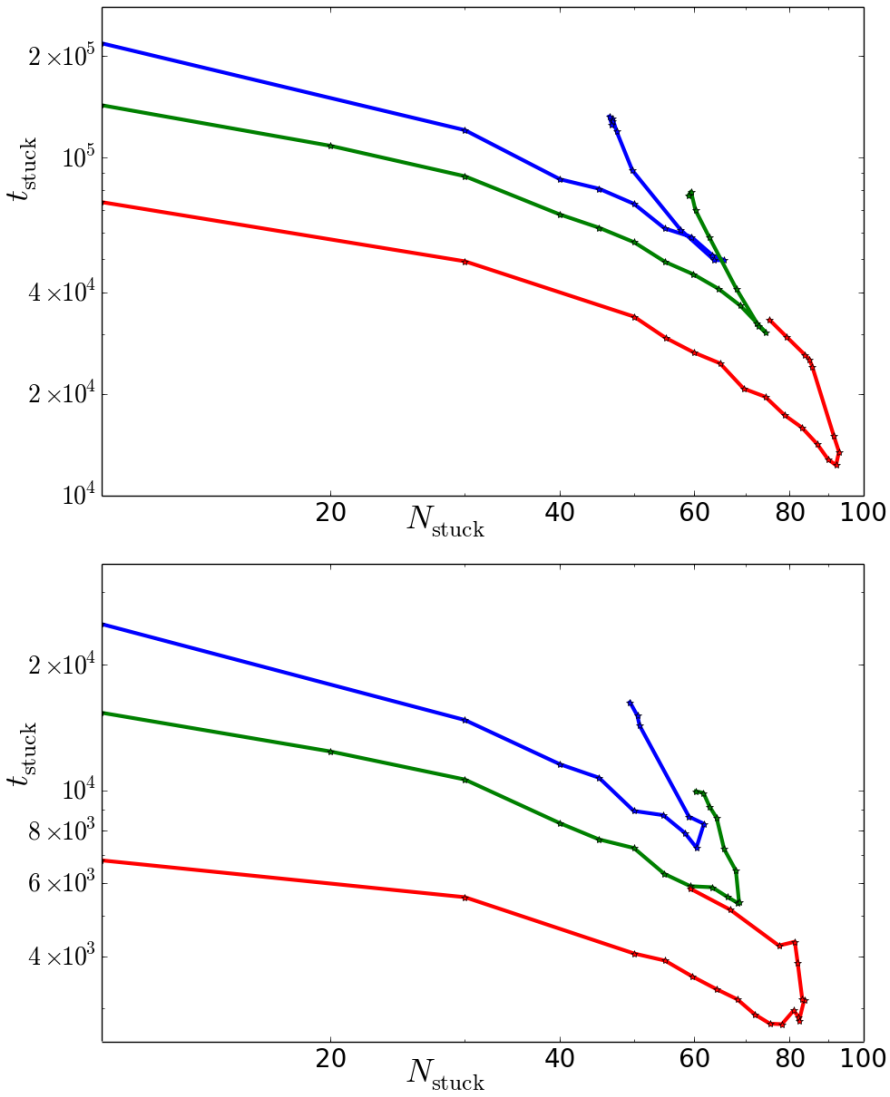

Figure 2.32: Plot of t_{stuck} against N_{stuck} for the cavity-nanopore with $F_{\mathrm{pore}} = 0.4\epsilon/\sigma$ (top, blue), $0.5\epsilon/\sigma$ (middle, green), and $0.75\epsilon/\sigma$ (bottom, red) and $r_{\mathrm{eff,cavi}} = 4.0\sigma$. The top plot corresponds to the case with the narrow pore, whereas the bottom is for the case with the wide pore.

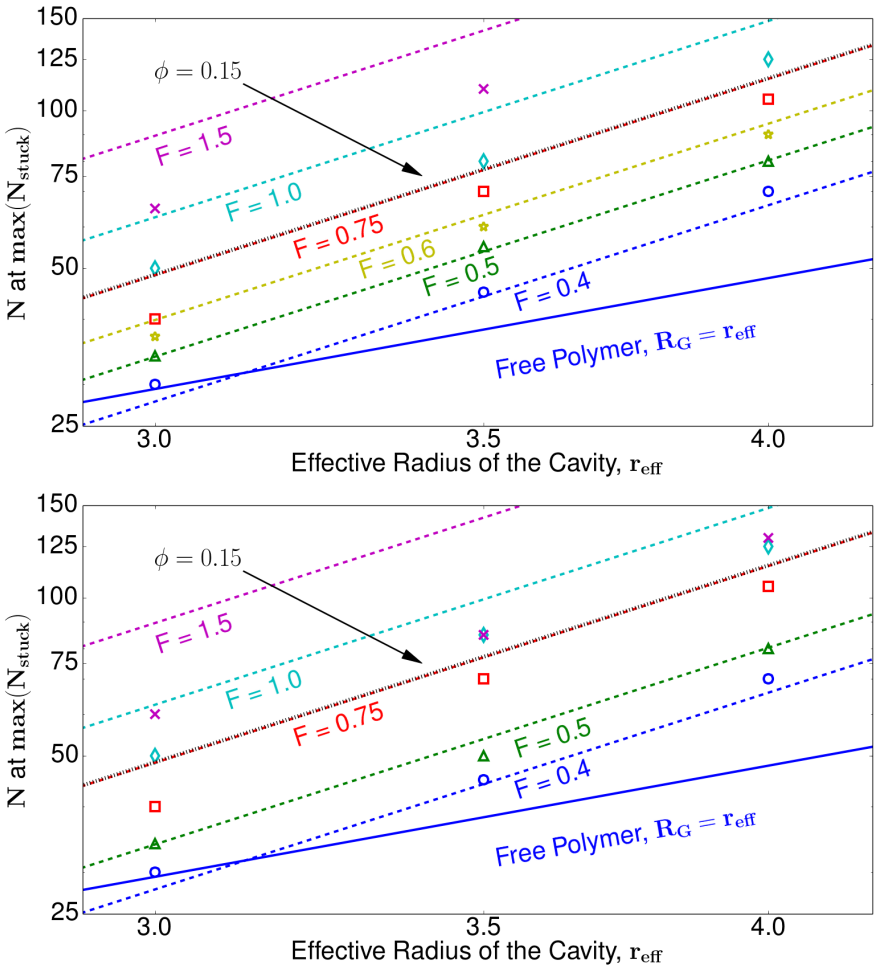

Figure 2.33: Comparison of the free energy model for $r_{\text{nom,pore}} = 1.3\sigma$ (top) and $r_{\text{nom,pore}} = 1.5\sigma$ (bottom) to the values of N measured from simulation to maximize N_{stuck}.

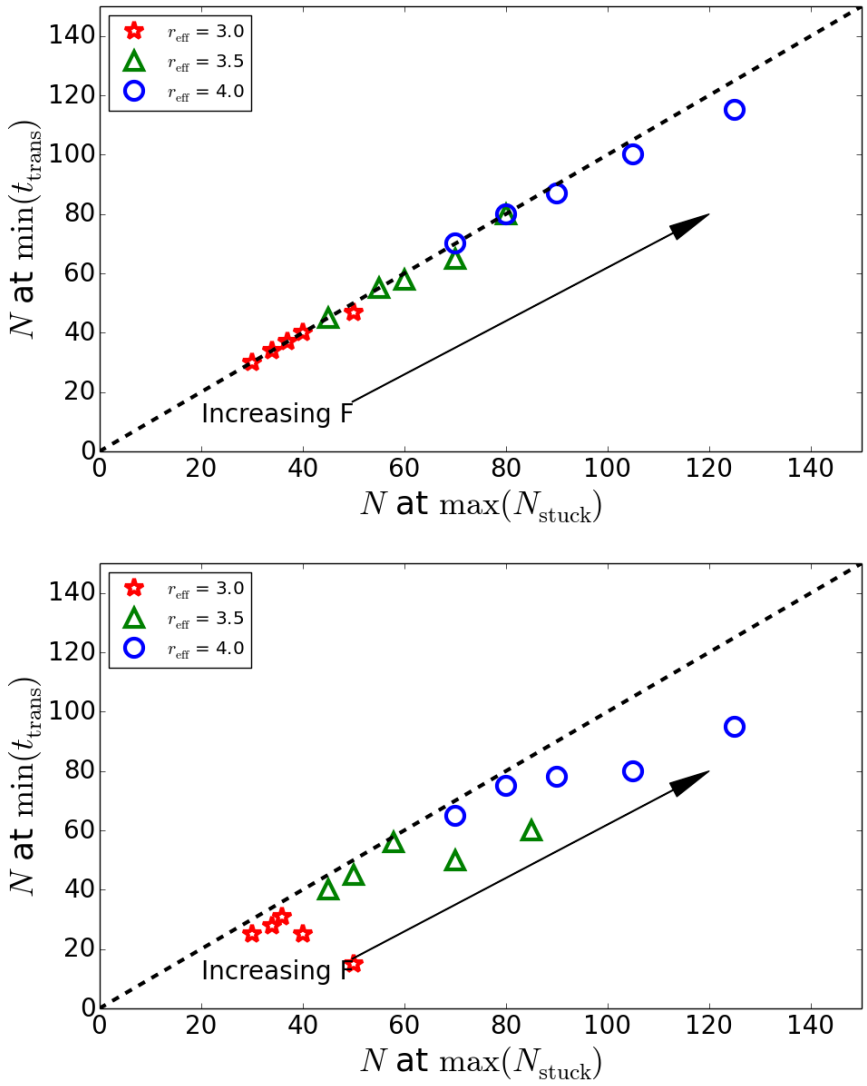

Figure 2.34: Comparison of N^*_{trans} and N^*_{stuck} for $r_{\text{eff,pore}} = 1.3\sigma$ (top) and $r_{\text{eff,pore}} = 1.5\sigma$ (bottom).

ing the low force, narrow pore regime for which Equation 2.41 was constructed. Since N^* is an important parameter for applications, it is worthwhile to understand why and how the predictions of Equation 2.41 differ from the true N^* when the system parameters are not in the narrow pore, low-F_{pore} regime. Figure 2.34 compares N^*, which minimizes t_{trans}, to the N that maximizes N_{stuck}, which is well predicted by Equation 2.41 even outside the narrow pore, low-F_{pore} regime.

At low forces, the two values are the same. As the force is increased, $N^* < \text{argmin}_N(N_{\text{stuck}})$. In other words, the chain length that translocates most quickly on average is somewhat shorter than the chain length that maximizes the average N_{stuck}. This implies that Equation 2.41 provides a robust *upper bound* for the true value of N^*_{trans} even away from the low force, narrow pore regime for which it was conceived.

The result that $N^* < \text{argmin}_N(N_{\text{stuck}})$ is not easy to explain. One possible mechanism for this behaviour might be as follows. The plots comparing t_{trans} or t_{stuck} over a range of applied forces (Figs 2.19 and 2.20) demonstrated that the magnitudes of t_{trans} and t_{stuck} diminish very rapidly as the applied force is increased. As such, as the force increases, non-equilibrium dynamics will naturally become more significant. Perhaps the result that $N^* < \text{argmin}_N(N_{\text{stuck}})$ at higher forces arises because the equilibrium picture of the stuck phase becomes progressively less valid as the force increases. Further work will be performed to investigate this possibility.

In summary, this section developed a model to predict the critical chain length that minimizes the translocation time t_{trans}. The model is also effective at predicting the chain length that maximizes N_{stuck}, the occupancy of the cavity during the stuck phase. These two quantities coincide closely with one another in the low force, narrow pore regime. Upper and lower bounds of applicability

for the model are well understood. Even when the model fails to predict the critical chain length, Equation 2.41 still seems to provide a reliable upper bound for N^*.

Chapter 3

Conclusion

In this thesis, the first attempts at characterizing polymer translocation through a novel cavity-nanopore geometry have been summarized. The translocation time was found to have a very pronounced minimum at an intermediate chain length, named the critical chain length. Chains that are either longer or shorter than the critical chain length take significantly longer to translocate. This qualitative result is of great interest for future application of the cavity-nanopore geometry.

Furthermore, a detailed characterization of the translocation process was used to explore the dynamics of the system over a range of simulation parameters. This uncovered the dramatic effect of tail formation on the translocation process. In particular, the formation of a tail in the *cis* region during the stuck phase for chains longer than the critical length is responsible for the extremely rapid growth of translocation time for $N \gtrsim N^*$. As a result of this mechanism, the minimum in the translocation time at the critical length is extremely pronounced. This has implications for applications, as discussed below.

Other dynamics were also explored. The behaviour of the exit time was ex-

plained using the tension-propagation model of polymer translocation. Varying cavity size and applied electric field by small amounts was found to rescale the behaviour of the system, preserving the qualitative results. Conversely, increasing the applied field strength by a large amount can lead to a new regime of operation, where translocation time is roughly constant for chains shorter than the critical chain length, but increases rapidly for chains longer than this. Finally, increasing the width of the entrance and exit pores was found to reduce the sharpness of the transition at the critical chain length.

The understanding of the system obtained from these analyses was combined into a simplified theoretical model of polymer translocation through the cavity-nanopore. The free energy of this model was described in the limit of narrow pores and low applied field strengths. This free energy was used to estimate the critical chain length in that regime. Good agreement was recovered between this theoretical prediction and the measurements made from simulations. Furthermore, bounds of applicability of the model were derived and shown to be in good agreement with the data. Finally, it was shown that, outside these bounds of applicability, the model still seems to provide a robust upper bound for N^*.

The application of nanopores for sorting of polymer mixtures by chain length was discussed in the Introduction. Forcing a polymer mixture through one or more nanopores of traditional design sorts chains by length from shortest to longest. This order arises because translocation time through standard nanopores is a monotonic function of chain length. In this sense, a standard nanopore acts as a low-pass filter for polymer length. Conversely, translocation time through a cavity-nanopore is a non-monotonic function of chain length with a sharp minimum. This suggests that the cavity-nanopore system should act as a band-pass filter for polymer length. Furthermore, since the minimum in translocation time is very sharp, the filtering action of the CN is also expected

to be highly selective.

The success of the free energy model in predicting the critical chain length also has implications for experimental applications of the CN. The corresponding equation for N^* captures the effect of the applied field strength on critical chain length. Using this relationship, the critical chain length of a single cavity-nanopore device could be dynamically tuned on-the-fly during applications. Conversely, the relationship between cavity size and critical chain length could be used to guide the construction of an optimal pore geometry for a given application.

As alluded to above, this thesis is a summary of what are only the first steps in the characterization of this novel device. The rich variety of physical phenomena and promising applications uncovered in this preliminary study certainly underscores the need for further study. With regards to future simulation work, the most promising avenues of research include:

- Incorporating a full electric field into the simulations, rather than the purely axial field used here.
- Specializing the polymer model to represent DNA molecules, which are the molecule of primary interest in nanopore applications.
- Modelling the capture process, whereby polymers first enter the nanopore from free solution in the *cis* region.
- Simulating the filtering action of cavity-nanopores placed in series, to directly evaluate its potential as a filter on polymer length.
- Simulation the translocation of molecules other than generic linear polymers through the cavity-nanopore.

Several of these projects are already being pursued.

Appendix A

Fail Rates

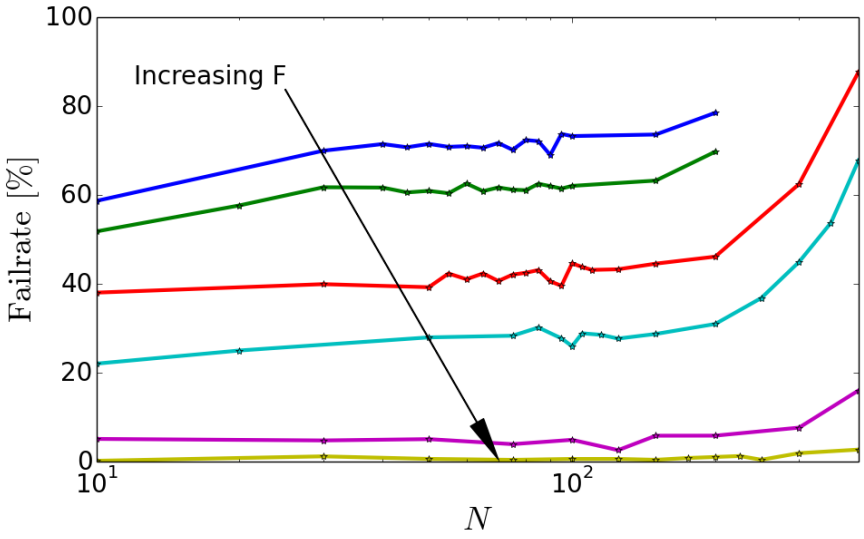

Figure A.1: Rate of failed events as a function of chain length for $r_{\text{nom,pore}} = 1.3\sigma$, $r_{\text{eff,cavi}} = 4.0\sigma$, and the range of forces from Figure 2.19.

Figure A.1 shows the rate of failed events for a sample of simulation param-

eters. The fail rate is computed as

$$\text{Fail Rate} = \frac{\text{Number of Failed Events}}{\text{Number of Successful Events} + \text{Number of Failed Events}}. \quad \text{(A.1)}$$

The fail rate increases as the force is reduced, suggesting that in practice there is a lower limit on the force at which the CN can be operated.

It is interesting to note that the non-monotonic behaviour of the translocation time for successful events is preserved regardless of the fail rate. In particular, consider the topmost fail rate plot of Figure A.1. This data corresponds to the basic regime described in the Results section. Thus, for $N < 75$ in this regime, the translocation time of successful events is decreasing with N even as the failure rate is increasing with N.

Bibliography

[1] *www.sharcnet.ca*.

[2] George Keith Batchelor. *An introduction to fluid dynamics*. Cambridge university press, 2000.

[3] SM Bezrukov. Ion channels as molecular coulter counters to probe metabolite transport. *Journal of Membrane Biology*, 174(1):1–13, 2000.

[4] Twist Bioscience. *Twist Bioscience Announces Microsoft Purchase of its Synthetic DNA for Digital Data Storage Research*. 2016.

[5] Boyang Cao, Rongrong Li, Songjin Xiong, Fangfang Yao, Xiangqian Liu, Min Wang, Lu Feng, and Lei Wang. Use of a dna microarray for detection and identification of bacterial pathogens associated with fishery products. *Applied and environmental microbiology*, 77(23):8219–8225, 2011.

[6] Jeffrey Chuang, Yacov Kantor, and Mehran Kardar. Anomalous dynamics of translocation. *Phys. Rev. E*, 65:011802, Dec 2001.

[7] Scott L Cockroft, John Chu, Manuel Amorin, and M Reza Ghadiri. A single-molecule nanopore device detects dna polymerase activity with single-nucleotide resolution. *Journal of the American Chemical Society*, 130(3):818–820, 2008.

[8] Hendrick W. de Haan and Gary W. Slater. Memory effects during the unbiased translocation of a polymer through a nanopore. *The Journal of Chemical Physics*, 136(15), 2012.

[9] Hendrick W. de Haan and Gary W. Slater. Using an incremental mean first passage approach to explore the viscosity dependent dynamics of the unbiased translocation of a polymer through a nanopore. *The Journal of Chemical Physics*, 136(20), 2012.

[10] J. L. A. Dubbeldam, A. Milchev, V. G. Rostiashvili, and T. A. Vilgis. Polymer translocation through a nanopore: A showcase of anomalous diffusion. *Phys. Rev. E*, 76:010801, Jul 2007.

[11] Olga K Dudko, Jérôme Mathé, Attila Szabo, Amit Meller, and Gerhard Hummer. Extracting kinetics from single-molecule force spectroscopy: nanopore unzipping of dna hairpins. *Biophysical journal*, 92(12):4188–4195, 2007.

[12] Christopher M. Edmonds, Yeny C. Hudiono, Amir G. Ahmadi, Peter J. Hesketh, and Sankar Nair. Polymer translocation in solid-state nanopores: Dependence of scaling behavior on pore dimensions and applied voltage. *The Journal of Chemical Physics*, 136(6), 2012.

[13] Yanxiao Feng, Yuechuan Zhang, Cuifeng Ying, Deqiang Wang, and Chunlei Du. Nanopore-based fourth-generation {DNA} sequencing technology. *Genomics, Proteomics & Bioinformatics*, 13(1):4 – 16, 2015.

[14] Daniel Fologea, James Uplinger, Brian Thomas, David S. McNabb, and Jiali Li. Slowing dna translocation in a solid-state nanopore. *Nano Letters*, 5(9):1734–1737, 2005. PMID: 16159215.

[15] Michel G. Gauthier and Gary W. Slater. A monte carlo algorithm to study polymer translocation through nanopores. ii. scaling laws. *The Journal of Chemical Physics*, 128(20), 2008.

[16] Carl P Goodrich, Serdal Kirmizialtin, Beatrice M Huyghues-Despointes, Aiping Zhu, J Martin Scholtz, Dmitrii E Makarov, and Liviu Movileanu. Single-molecule electrophoresis of β-hairpin peptides by electrical recordings and langevin dynamics simulations. *The Journal of Physical Chemistry B*, 111(13):3332–3335, 2007.

[17] G. S. Grest and K. Kremer. Molecular dynamics simulation for polymers in the presence of a heat bath. *Physical Review, A*, 33(3628), 1986.

[18] Li-Qun Gu, Orit Braha, Sean Conlan, Stephen Cheley, and Hagan Bayley. Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature*, 398(6729):686–690, 1999.

[19] Li-Qun Gu, Meni Wanunu, Michael X Wang, Larry McReynolds, and Yong Wang. Detection of mirnas with a nanopore single-molecule counter. *Expert review of molecular diagnostics*, 12(6):573–584, 2012.

[20] C.W. H. Means for counting particles suspended in a fluid, October 20 1953. US Patent 2,656,508.

[21] Zachary D. Harms, Klaus B. Mogensen, Pedro S. Nunes, Kaimeng Zhou, Brett W. Hildenbrand, Indranil Mitra, Zhenning Tan, Adam Zlotnick, Jrg P. Kutter, and Stephen C. Jacobson. Nanofluidic devices with two pores in series for resistive-pulse sensing of single virus capsids. *Analytical Chemistry*, 83(24):9573–9578, 2011. PMID: 22029283.

[22] Breton Hornblower, Amy Coombs, Richard D Whitaker, Anatoly Kolomeisky, Stephen J Picone, Amit Meller, and Mark Akeson. Single-

molecule analysis of dna-protein complexes using nanopores. *Nature Methods*, 4(4):315–317, 2007.

[23] Stefan Howorka, Liviu Movileanu, Orit Braha, and Hagan Bayley. Kinetics of duplex formation for individual dna strands within a single protein nanopore. *Proceedings of the National Academy of Sciences*, 98(23):12996–13001, 2001.

[24] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[25] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[26] Ilkka Huopaniemi, Kaifu Luo, Tapio Ala-Nissila, and See-Chen Ying. Langevin dynamics simulations of polymer translocation through nanopores. *The Journal of Chemical Physics*, 125(12), 2006.

[27] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, and W. Sung. Influence of non-universal effects on dynamical scaling in driven polymer translocation. *The Journal of Chemical Physics*, 137(8), 2012.

[28] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, and W. Sung. Unifying model of driven polymer translocation. *Phys. Rev. E*, 85:051803, May 2012.

[29] Xiao-feng Kang, Stephen Cheley, Xiyun Guan, and Hagan Bayley. Stochastic detection of enantiomers. *Journal of the American Chemical Society*, 128(33):10684–10685, 2006.

[30] Yacov Kantor and Mehran Kardar. Anomalous dynamics of forced translocation. *Phys. Rev. E*, 69:021806, Feb 2004.

[31] Felix Kapahnke, Ulrich Schmidt, Dieter W. Heermann, and Matthias Weiss. Polymer translocation through a nanopore: The effect of solvent conditions. *The Journal of Chemical Physics*, 132(16), 2010.

[32] Niki Karachaliou, Clara Mayo-de-las Casas, Miguel Angel Molina-Vila, and Rafael Rosell. Real-time liquid biopsies become a reality in cancer treatment. *Annals of translational medicine*, 3(3), 2015.

[33] JohnJ. Kasianowicz, Eric Brandin, Daniel Branton, and DavidW. Deamer. Characterization of individual polynucleotide molecules using a membranechannel. *Proceedings of the National Academy of Sciences*, 93(24):13770–13773, 1996.

[34] Sara Huston Katsanis and Nicholas Katsanis. Molecular genetic testing and the future of clinical genomics. *Nature Reviews Genetics*, 14(6):415–426, 2013.

[35] Stefan Kesselheim, Wojciech Müller, and Christian Holm. Origin of current blockades in nanopore translocation experiments. *Physical review letters*, 112(1):018101, 2014.

[36] Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[37] Harold Kwok, Kyle Briggs, and Vincent Tabard-Cossa. Nanopore fabrication by controlled dielectric breakdown. *PLoS ONE*, 9(3):1–6, 03 2014.

[38] Martin Langecker, Daniel Pedone, Friedrich C. Simmel, and Ulrich Rant. Electrophoretic time-of-flight measurements of single dna molecules with two stacked nanopores. *Nano Letters*, 11(11):5002–5007, 2011.

[39] T Laver, J Harrison, PA ONeill, K Moore, A Farbos, K Paszkiewicz, and David J Studholme. Assessing the performance of the oxford nanopore technologies minion. *Biomolecular detection and quantification*, 3:1–8, 2015.

[40] Kate R Lieberman, Gerald M Cherf, Michael J Doody, Felix Olasagasti, Yvette Kolodji, and Mark Akeson. Processive replication of single dna molecules in a nanopore catalyzed by phi29 dna polymerase. *Journal of the American Chemical Society*, 132(50):17961–17972, 2010.

[41] H. J. Limbach, A. Arnold, B. A. Mann, and C. Holm. ESPResSo – an extensible simulation package for research on soft matter systems. *Comp. Phys. Comm.*, 174(9):704–727, May 2006.

[42] Jianxun Lin, Anatoly Kolomeisky, and Amit Meller. Helix-coil kinetics of individual polyadenylic acid molecules in a protein channel. *Physical review letters*, 104(15):158101, 2010.

[43] Xu Liu, Mirna Mihovilovic Skanata, and Derek Stein. Entropic cages for trapping dna near a nanopore. *Nat Commun*, 6, Feb 2015. Article.

[44] David K. Lubensky and David R. Nelson. Driven polymer translocation through a narrow pore. *Biophysical Journal*, 77(4):1824 – 1838, 1999.

[45] Kaifu Luo, T. Ala-Nissila, and See-Chen Ying. Polymer translocation through a nanopore: A two-dimensional monte carlo study. *The Journal of Chemical Physics*, 124(3), 2006.

[46] Martin Magill, Cory Falconer, Ed Waller, and Hendrick W de Haan. Translocation time through a nanopore with an internal cavity is minimal for polymers of intermediate length. *Physical Review Letters*, 117(24):247802, 2016.

[47] Harri Mökkönen, Timo Ikonen, Hannes Jnsson, and Tapio Ala-Nissila. Polymer escape from a confining potential. *The Journal of Chemical Physics*, 140(5), 2014.

[48] M. Muthukumar. Polymer translocation through a hole. *The Journal of Chemical Physics*, 111(22):10371–10374, 1999.

[49] Vladimir V. Palyulin, Tapio Ala-Nissila, and Ralf Metzler. Polymer translocation: the first two decades and the recent diversification. *Soft Matter*, 10:9016–9037, 2014.

[50] Debabrata Panja. Response of single polymers to localized step strains. *Phys. Rev. E*, 79:011803, Jan 2009.

[51] Debabrata Panja, Gerard T Barkema, and Robin Ball. Origin and effects of anomalous dynamics on unbiased polymer translocation. 2007.

[52] Debabrata Panja, Gerard T Barkema, and Robin C Ball. Anomalous dynamics of unbiased polymer translocation through a narrow pore. *Journal of Physics: Condensed Matter*, 19(43):432202, 2007.

[53] Debabrata Panja, Gerard T Barkema, and Robin C Ball. Polymer translocation out of planar confinements. *Journal of Physics: Condensed Matter*, 20(7):075101, 2008.

[54] Debabrata Panja, Gerard T Barkema, and Anatoly B Kolomeisky. Through the eye of the needle: recent advances in understanding biopolymer translocation. *Journal of Physics: Condensed Matter*, 25(41):413101, 2013.

[55] Daniel Pedone, Martin Langecker, Gerhard Abstreiter, and Ulrich Rant. A porecavitypore device to trap and investigate single nanoparticles and dna molecules in a femtoliter compartment: Confined diffusion and narrow escape. *Nano Letters*, 11(4):1561–1567, 2011. PMID: 21388205.

[56] Joseph WF Robertson, Claudio G Rodrigues, Vincent M Stanford, Kenneth A Rubinson, Oleg V Krasilnikov, and John J Kasianowicz. Single-molecule mass spectrometry in solution using a solitary nanopore. *Proceedings of the National Academy of Sciences*, 104(20):8207–8211, 2007.

[57] Lutz Roewer. Dna fingerprinting in forensics: past, present, future. *Investigative genetics*, 4(1):1, 2013.

[58] Ankit Rohatgi. Webplotdigitizer. *URL http://arohatgi.info/WebPlotDigitizer/app*, 2011.

[59] T. Saito and T. Sakaue. Dynamical diagram and scaling in polymer driven translocation. *The European Physical Journal E*, 34(12):1–8, 2011.

[60] Takuya Saito and Takahiro Sakaue. Erratum to: Dynamical diagram and scaling in polymer driven translocation. *The European Physical Journal E*, 35(11):1–2, 2012.

[61] Takahiro Sakaue. Nonequilibrium dynamics of polymer translocation and straightening. *Phys. Rev. E*, 76:021803, Aug 2007.

[62] Takahiro Sakaue. Sucking genes into pores: Insight into driven translocation. *Phys. Rev. E*, 81:041808, Apr 2010.

[63] Takahiro Sakaue, Takuya Saito, and Hirofumi Wada. Dragging a polymer in a viscous fluid: Steady state and transient. *Phys. Rev. E*, 86:011804, Jul 2012.

[64] Alexis F Sauer-Budge, Jacqueline A Nyamwanda, David K Lubensky, and Daniel Branton. Unzipping kinetics of double-stranded dna in a nanopore. *Physical Review Letters*, 90(23):238101, 2003.

[65] Tamar Schlick. *Molecular modeling and simulation: an interdisciplinary guide: an interdisciplinary guide*, volume 21. Springer Science & Business Media, 2010.

[66] David Sean, Hendrick W. de Haan, and Gary W. Slater. Translocation of a polymer through a nanopore starting from a confining nanotube. *Electrophoresis*, 36(5):682–691, 2015.

[67] Gary W Slater, Claude Desruisseaux, Sylvain J Hubert, Jean-François Mercier, Josée Labrie, Justin Boileau, Frédéric Tessier, and Marc P Pépin. Theory of dna electrophoresis: A look at some current challenges. *Electrophoresis*, 21(18):3873–3887, 2000.

[68] Ralph M. M. Smeets, Ulrich F. Keyser, Diego Krapf, Meng-Yue Wu, Nynke H. Dekker, and Cees Dekker. Salt dependence of ion transport and dna translocation through solid-state nanopores. *Nano Letters*, 6(1):89–95, 2006.

[69] L. Song, M. Hobaugh, C. Shustak, S. Cheley, H. Bayley, and J.E. Gouaux. alpha-hemolysin heptamer. *Wikimedia Commons*, 12-05-2016. File: 7ahl.jpg.

[70] Christopher C Striemer, Thomas R Gaborski, James L McGrath, and Philippe M Fauchet. Charge-and size-based separation of macromolecules using ultrathin silicon membranes. *Nature*, 445(7129):749–753, 2007.

[71] Li-Zhen Sun and Meng-Bo Luo. Langevin dynamics simulation on the translocation of polymer through alpha-hemolysin pore. *Journal of Physics: Condensed Matter*, 26(41):415101, 2014.

[72] W. Sung and P. J. Park. Polymer translocation through a pore in a membrane. *Phys. Rev. Lett.*, 77:783–786, Jul 1996.

[73] Iwao Teraoka. *Polymer Solutions*. John Wiley & Sons, Inc., 2002.

[74] Wenonah Vercoutere, Stephen Winters-Hilt, Hugh Olsen, David Deamer, David Haussler, and Mark Akeson. Rapid discrimination among individual dna hairpin molecules at single-nucleotide resolution using an ion channel. *Nature biotechnology*, 19(3):248–252, 2001.

[75] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, Jul 1967.

[76] Dongshan Wei, Wen Yang, Xigao Jin, and Qi Liao. Unforced translocation of a polymer chain through a nanopore: The solvent effect. *The Journal of Chemical Physics*, 126(20), 2007.

[77] Aaron J Wolfe, Mohammad M Mohammad, Stephen Cheley, Hagan Bayley, and Liviu Movileanu. Catalyzing the translocation of polypeptides through attractive interactions. *Journal of the American Chemical Society*, 129(45):14034–14041, 2007.