MCCXXII

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”
Corso di Laurea in Fisica

Tesi di Laurea

Processi di diffusione e traslocazione dei polimeri

Polymers diffusion and translocation

Relatore
Prof. Fulvio Baldovin

Laureando
Giuseppe Rizzi

Anno Accademico 2023/2024

Riassunto

Recenti sviluppi nell'analisi dei processi diffusivi in ambienti complessi, come la diffusione di un polimero in una cellula, mostrano che la funzione di densità di probabilità della coordinata diffusiva manifesta caratteristiche non gaussiane. Tali processi possono essere modellizzati facendo fluttuare nel tempo il coefficiente di diffusione, che allora diventa una seconda sorgente di aleatorietà, in aggiunta alla coordinata diffusiva. In teoria della probabilità si parla di subordinazione, visto che la coordinata diffusiva è subordinata a quella del coefficiente di diffusione. Inoltre, nello studio del comportamento dei polimeri in ambienti cellulari, la traslocazione gioca un ruolo fondamentale. Nella sintesi delle proteine, ad esempio, l'RNA messaggero deve trasportare l'informazione genetica fuori dal nucleo, e per farlo deve attraversare un poro sulla superficie nucleare, deve cioè traslocare. Nel corso della tesi, dopo aver rivisto le equazioni di Langevin, esploreremo la diffusione dei polimeri sia in condizioni standard (diffusione browniana) che in presenza di altri monomeri (diffusione browniana non gaussiana); ed infine tratteremo il caso della traslocazione da un punto di vista termodinamico e stocastico.

Abstract

Recent trends in the analysis of diffusion processes in complex environments, such as the diffusion of a polymer in a crowded biological cell, focus on non-Gaussian traits for the probability density function of the diffusing coordinate. One possibility of modelling such processes is by allowing the diffusion coefficient to fluctuate in time, as a consequence of a second source of randomness, besides the one affecting the diffusing coordinate. In terms of probability theory, this mechanism is called subordination, since the randomness of the diffusing coordinate is subordinated to that of the diffusion coefficient. Additionally, the ability of polymers to undergo significant conformational changes makes them ideal candidates for large entropic variations. This characteristic trait will be employed in the analysis of the traslocation process, which, along with simple diffusion, governs much of the polymer dynamics in biological cells. In fact, it is a paramount biological paradigm for the transmission of information at the cellular level. In this thesis, after reviewing the Langevin equations of motion, we will explore the diffusion of polymers not only under standard conditions (Brownian diffusion), but also in a medium full of other monomers (Brownian non-Gaussian diffusion). Finally, the translocation process will be illustrated by means of thermodinamic and stochastic tools.

Contents

Chapter 1

A general overview of Brownian motion

The botanist Robert Brown was the first to observe under the microscope the dispersion of pollen grains (diameter $\approx 1\,\mu$m) in water, whose chaotic motion was later given the name Brownian motion. This term actually refers to the movement of any colloidal particle, typically with a diameter ranging from a few nanometres to about $1\,\mu$m, in a dispersive medium. When dispersed, these particles do not dissolve or mix with this substance but instead retain their chemical structure. The description of the phenomenon was left unsolved until the landmark paper of Einstein in 1905. Einstein realised, among other things, that when the collective effect of many interactions occurs on a given particle, Newton's law of motion needs to be stochastically generalised to take into account random variables. At this scale, physics inherently demands a stochastic approach rather than a deterministic one; indeed, physical quantities such as the position and velocity of these colloidal particles will be described with probability density functions instead of ordinary deterministic ones. In this chapter, we will establish the foundation for understanding the mathematics behind Brownian motion by first considering the *random walk* model for discrete steps and then the *Langevin equation* for a continuous description. Consequently, we will extend the discussion to polymers - long chains made up of identical particles called monomers - to examine their dynamics in a generic medium, by introducing the *Rouse model*.

1.1 The diffusion law

Before giving a detailed description of Brownian motion, we first examine a mathematical model known as *random walk* which gives a first idea of the problem under study; in fact, its variance, referred to as the *diffusion law*, is identical to that of colloidal particles in a medium under an overdamped regime (as we shall see in the next chapter).

The random walk is built up by independent steps of equal length $\Delta L > 0$, which occur at regular time intervals of span $\Delta t = t/N$ with N the total number of steps made. For simplicity, we consider a one-dimensional walk starting from $X(0) = x_0$ at time $t = 0$ with equal probability of moving in the two allowed directions (which we call from now on *left* and *right*).

Let us call $X(N)$ the position of the walker (we can think of a particle) after N steps; on average, its value must be equal to the starting point for the equiprobability of the steps, thus

$$\mathbb{E}[X(N)] = x_0 \quad \forall N \geq 0. \tag{1.1}$$

We now introduce the random variable $K(N)$ which can assume only two values

$$k(N) = \pm\Delta L \quad \forall N \geq 1, \tag{1.2}$$

and its probability mass function can be written as

$$P_K(k) = \begin{cases} 1/2 & \text{if } k = \Delta L \quad \text{right step,} \\ 1/2 & \text{if } k = -\Delta L \quad \text{left step.} \end{cases}$$

This variable represents the random displacement associated with the N-th step. Using $\mathbb{E}[K^m] \equiv \sum_k k^m \, P_K(k)$, it is easy to find the following expressions for the mean and variance of the process $K(N)$:

$$\mathbb{E}[K(N)] = (+\Delta L)\frac{1}{2} + (-\Delta L)\frac{1}{2} = 0, \tag{1.3}$$

$$\mathbb{E}[K(N)^2] = (+\Delta L)^2\frac{1}{2} + (-\Delta L)^2\frac{1}{2} = (\Delta L)^2. \tag{1.4}$$

By construction, $X(N)$ can be broken down as follows:

$$X(N) = X(N-1) + K(N), \tag{1.5}$$

which allows us to find the variance of the *random walk*,

$$\begin{aligned}
\mathrm{VAR}[X(N)] &\equiv \mathbb{E}[X(N)^2] - \mathbb{E}[X(N)]^2 \\
&= \mathbb{E}[(X(N-1) + K(N))^2] - x_0^2 \\
&= \mathbb{E}[X(N-1)^2] + 2\mathbb{E}[X(N-1)\,K(N)] + \mathbb{E}[K(N)^2] - x_0^2 \\
&= (\text{independent steps imply } \mathrm{COV}[X(N-1), K(N)] = 0)\,^{1} \\
&= \mathbb{E}[X(N-1)^2] + 2\mathbb{E}[X(N-1)]\mathbb{E}[K(N)] + \mathbb{E}[K(N)^2] - x_0^2 \\
&= (\text{recall Eqs. (1.3) and (1.4)}) \\
&= \mathbb{E}[X(N-1)^2] + \Delta L^2 - x_0^2. \qquad (1.6)
\end{aligned}$$

This expression yields:

$$\mathbb{E}[X(N)^2] = \mathbb{E}[X(N-1)^2] + \Delta L^2 \quad \text{for all } N \geq 0, \tag{1.7}$$

therefore, for $N-1$ it reads: $\mathbb{E}[(X(N-1)^2] = \mathbb{E}[X(N-2)^2] + \Delta L^2$. We substitute this result into Eq. (1.7) and obtain $\mathbb{E}[X(N)^2] = \mathbb{E}[X(N-2)^2] + 2\Delta L^2$. By repeating this process N times we end up with:

$$\begin{aligned}
\mathbb{E}[X(N)^2] &= \mathbb{E}[(X(N-N))^2] + N\Delta L^2 \qquad (1.8) \\
&= x_0^2 + N\Delta L^2, \qquad (1.9)
\end{aligned}$$

thus the variance of $X(N)$ is:

$$\mathrm{VAR}[X(N)] = N\Delta L^2. \tag{1.10}$$

Recalling that $\Delta t = t/N \Rightarrow N = t/\Delta t$ and introducing the *diffusion coefficient*:

$$D \equiv \frac{\Delta L^2}{2\Delta t}, \tag{1.11}$$

we get the *diffusion law*

$$\mathrm{VAR}[X(t)] = 2D\,t. \tag{1.12}$$

If, instead of considering a one-dimensional random walk, we consider a model in which the steps all occur simultaneously and independently along three perpendicular directions, the diffusion law generalizes as follows: $\mathbb{E}[\mathbf{R}(t)^2] = \mathbb{E}[X(t)^2] + \mathbb{E}[Y(t)^2] + \mathbb{E}[Z(t)^2] = 6D\,t$; such that:

$$\mathrm{VAR}[\mathbf{R}(t)] = 6D\,t. \tag{1.13}$$

We just mention here that the diffusion law is of a universal nature. In fact, if we suppose that the steps occur with a length multiple of $\Delta L > 0$, provided that they are independent and identically distributed, the diffusion law remains the same.

---

[1] where, $\mathrm{COV}[X(N-1), K(N)] \equiv \mathbb{E}[X(N-1)K(N)] - \mathbb{E}[X(N-1)]\mathbb{E}[K(N)]$ measures the joint variability of two random variables.

1.2 The work of Langevin

We want to give a very concise, yet complete description of the phenomenon of Brownian motion by using the approach of Langevin. The mathematical model for Brownian motion that works with continuously distributed displacements in continuous time is indeed called the *Langevin equation*. The idea of Langevin is to model the motion of dispersed colloidal particles by resolving the net force acting on them into three components:

1. a *dissipative drag force* $-\gamma\,\mathbf{V}(t)$ proportional to the particle's velocity where $\gamma = 6\pi\mu R$ [2];
2. a *zero-mean, temporally uncorrelated randomly fluctuating force* $\mathbf{F}(t)$, Gaussian distributed;
3. an *external constant conservative force field* $-\nabla\Phi(\mathbf{r})$ (that will be ignored for the moment).

Physically speaking, we can imagine a rigid spherical particle that, having a radius much greater than the size of a single solvent molecule, exhibits a restless motion. Indeed, due to thermal energy, the solvent molecules collectively impact against the particle from all directions in a random way, exerting a fluctuating force $\mathbf{F}(t)$. On the other hand, as the particle moves it will collide on average more often on the front side, thus it will also experience a deterministic force proportional to its velocity yet in the opposite direction $-\gamma\,\mathbf{V}(t)$.

By the definition of position and by virtue of the Newton's second law of motion (stochastically generalised to account for random variables), we can write the Langevin equations for a colloidal particle suspended in a generic medium (for simplicity we just consider the x direction):

$$\frac{dX(t)}{dt} = V(t), \tag{1.14}$$

$$\frac{dV(t)}{dt} = -\frac{\gamma\,V(t)}{m} + \frac{F(t)}{m}, \tag{1.15}$$

where both $X(t)$ and $V(t)$ (particle position and velocity) are stochastic processes i.e. random variables whose density functions depend parametrically on time t [4], and m is the mass of the colloidal particle. We will not provide the complete solution to this process (which is available, e.g, in [4]), instead we will focus on understanding how Brownian motion arises from these equations, its probability density function and its main characteristics. First of all, let us analyse the fluctuating force $F(t)$. One can show [4] that the fluctuating force can be written as:

$$F(t) = \gamma\,\sqrt{2D}\,\Gamma(t), \tag{1.16}$$

where D is the diffusion coefficient as defined in Eq. (1.11) and $\Gamma(t)$ is the *Gaussian white noise* process: $\Gamma(t) \equiv \lim_{dt\to 0} \mathbb{N}(0, 1/dt)$[3]. From the definition just made one easily finds:

$$\mathbb{E}[F(t)] = \mathbb{E}[\Gamma(t)] = 0, \tag{1.17}$$

$$\mathbb{E}[F(t')\,F(t'')] = \gamma^2\,2D\,\mathbb{E}[\Gamma(t')\,\Gamma(t'')] = \gamma^2\,2D\,\delta(t'-t''). \tag{1.18}$$

These two expressions are physically meaningful, given the symmetry requirements for the former, and the independence of the steps for the latter. Equation (1.18) is one form of the *flucutaction-dissipation theorem* [4].

---

[2] This is the Stokes friction law: it represents the drag coefficient provided that the particle is modelled as a sphere of radius R moving at velocity $\mathbf{V}(t)$ in a quiescent fluid with μ being its viscosity.

[3] $\mathbb{N}(\mu,\sigma^2)$ is the Gaussian random variable with mean μ and variance σ^2, whose associated probability density function is thus $P_{\mathbb{N}(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

1.2.1 The overdamped Langevin equation

Brownian motion arises when the *Langevin equations* are considered in the context of an overdamped regime. This can be crudely done by putting the instantaneous acceleration to zero [1]:

$$0 = -\frac{\gamma V(t)}{m} + \frac{F(t)}{m} \quad \Rightarrow \quad V(t) = \frac{F(t)}{\gamma}. \tag{1.19}$$

The reason for this approach follows from considering that an overdamped regime emerges during a time span of the order of m/γ which is called the Smoluchowski time scale. In fact, on this time scale - which is the time $\mathbf{V}(t)$ needs to lose memory of its initial value [1] - the typical distance travelled by a particle with respect to its radius R, provided that its initial velocity by equipartition is $\mathbf{v}_0 \approx \sqrt{3k_BT/m}$, reads

$$\frac{\mathbf{v}_0 \frac{m}{\gamma}}{R} = \frac{\sqrt{3k_BTm}}{6\pi\mu R^2} \approx 10^{-5}, \tag{1.20}$$

where 10^{-5} refers to normal colloidal particles [1]. Thus the particles hardly move on this time scale. Furthermore, from Eq. (1.19) we find for the position of the colloidal particle:

$$\frac{dX(t)}{dt} = \frac{F(t)}{\gamma}. \tag{1.21}$$

which is called *overdamped Langevin equation.*

Now, although the process $X(t)$ in (1.21) is written in a differential form it turns out to be nowhere differentiable (yet it is continuous [4]); in fact, stochastic differential equations are mathematically defined as update formulas. Consequently, it can be shown [4] that the most accurate way of represent the *overdamped Langevin equation* written above, is the following

$$X(t+dt) = X(t) + \sqrt{2D}\, dW(dt), \tag{1.22}$$

where $dW(dt)$ is called *Wiener process* defined as $dW(dt) \equiv \mathbb{N}(0, dt)$. The stochastic process $X(t)$ that follows the set of Eqs. (1.21) and (1.22) is called Brownian motion (or *Wiener process*) with diffusive constant D. We have therefore, accomplished the first objective outlined at the start of this section which was understanding how Brownian motion emerges from the context of diffusion; next, we want to specify its probability density function and analyse its main characteristics.

First of all let us show that the process $X(t)$ is normal by making use of the update formula (1.22). Given the initial condition $X(t_0) = \mathbb{N}(x_0, 0)$ (since at the beginning of diffusion, the Brownian particle is certainly at a specific position), Eq. (1.22) becomes: $X(t_0 + dt) = \mathbb{N}(x_0, 0) + \sqrt{2D}\,\mathbb{N}(0, dt)$. The latter expression is a linear combination of statistically independent normal random variables, and thus $X(t_0 + dt)$ must be itself normal (see [4]). If we now update the value of the stochastic variable $X(t_0 + dt)$ by another dt we get: $X((t_0 + dt) + dt) = X(t_0 + dt) + \sqrt{2D}\,\mathbb{N}(0, dt)$ which again is a linear combination of statistically independent normal random variables, so that $X(t_0 + 2dt)$ is itself normal. By induction, we infer that the Brownian motion $X(t)$ is normal for all $t \geq t_0$, therefore we only need to determine its mean and variance to fully describe it.

We can now compute the exact solution to the equations of Brownian motion. Let us first integrate both side of the *overdamped Langevin equation* (1.21) over the time interval [0, t]:

$$X(t) = x_0 + \frac{1}{\gamma}\int_0^t dt'\, F(t'). \tag{1.23}$$

By taking the average of this result and using Eq. (1.17), we find the mean of $X(t)$:

$$\mathbb{E}[X(t)] = x_0, \tag{1.24}$$

We now proceed to calculate its variance:

$$\begin{aligned}\mathrm{VAR}[X(t)] &\equiv \mathbb{E}\left[(X(t)-x_0)^2\right]\\ &= \frac{1}{\gamma^2}\int_0^t dt' \int_0^t dt'' \,\mathbb{E}\left[F(t')\,F(t'')\right]\\ &= (\text{recall Eq. (1.18)})\\ &= 2D\int_0^t dt' \int_0^t dt''\,\delta(t'-t'')\\ &= 2D\,t, \end{aligned} \tag{1.25}$$

which corresponds to the diffusion law given in Eq. (1.12). In summary, a Brownian motion with diffusive constant D is a stochastic process $X(t)$ of the form

$$X(t) = \mathbb{N}\left(x_0,\, 2D\,t\right). \tag{1.26}$$

This motion is generally referred to as *Brownian Gaussian* diffusion, where *Brownian* indicates the linear dependence on time of the mean-squared displacement, while *Gaussian* the PDF of the process being a normal distribution.

1.3 Polymer Dynamics

Up to this point, we have examined the diffusion of colloidal particles in a generic medium. Let us now address the dynamics of essential biological structures called polymers. Polymers are huge molecules constructed out of many identical building blocks called monomers which can vary from a few hundreds to many thousands and can assume an indefinitely large number of shapes (however, throughout the thesis we will ignore their chemical structure in favor of a general physical analysis). In this section we want to outline one of the simplest mathematical models that describes the dynamics of polymers when dispersed in a generic medium (for instance water), which is called the *Rouse model*. Specifically, we will track the center of mass of a polymer, focusing on the equation of motion it follows and the diffusion coefficient that can be associated to it. The Rouse model precisely addresses these questions by considering the polymer as a chain of monomers connected to one another by ideal springs. Of course, the motion of a monomer through the liquid will induce a velocity field in the medium which will be felt by all the others; however the model completely neglects this hydrodynamic effect.

1.3.1 The Rouse model

Here, we fully discuss the *Rouse mode*l of polymer dynamics using the *overdamped Langevin equation* given in Eq. (1.21). In this case, however, the monomers that make up the polymer do not only feel a fluctuating random force - resulting from the collisions with water molecules - but also a deterministic harmonic force due to the effect of the ideal springs. Thus, we need to account for the external conservative field $-\nabla\Phi(\mathbf{r})$ and, in this section, also extend our analysis to three dimensions. Eq. (1.21) becomes

$$\frac{d\mathbf{R}(t)}{dt} = -\frac{\nabla\Phi(\mathbf{r})}{\gamma} + \mathbf{f}(t), \tag{1.27}$$

where $\mathbf{f}(t) = \mathbf{F}(t)/\gamma$. We begin by considering a polymer chain composed of N identical monomers (each with a diffusive constant D) that we start counting from 1 (see Figure (1.1)). Since all monomers have the same mass m we can define the centre of mass of the polymer as follows

$$\mathbf{R}_{CM}(t) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{R}_n(t). \tag{1.28}$$

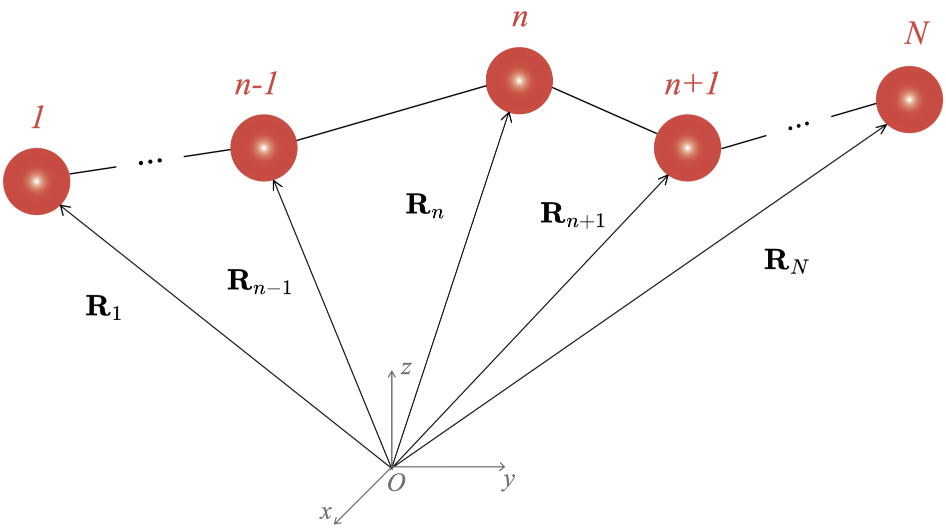

Figure 1.1: A polymer chain according to the Rouse model.

We now need to determine the equation of motion for the monomers. We examine the motion of one particular monomer, by fixing the positions of the other $N-1$, and we find that its velocity (in the overdamped regime) must adhere to Equation (1.27), with the external force being harmonic. For $n = 1$ and $n = N$ we simply find

$$\frac{d\mathbf{R}_1}{dt} = -k(\mathbf{R}_1 - \mathbf{R}_2) + \mathbf{f}_1 \quad \text{and} \quad \frac{d\mathbf{R}_N}{dt} = -k(\mathbf{R}_N - \mathbf{R}_{N-1}) + \mathbf{f}_N \tag{1.29}$$

where k is a constant that depends on the characteristics of both the medium and the monomer[4]. For the the n-th monomer we need to consider the spring attraction of the $(n-1)$-th and $(n+1)$-th monomers, thus

$$\frac{d\mathbf{R}_n}{dt} = -k[(\mathbf{R}_n - \mathbf{R}_{n+1}) + (\mathbf{R}_n - \mathbf{R}_{n-1})] + \mathbf{f}_n. \tag{1.30}$$

Let us now consider the *fluctuation-dissipation theorem* given in Eq. (1.18). First of all, we rewrite it in terms of $\mathbf{f}(t)$ such that it becomes: $\mathbb{E}[\mathbf{f}(t)\,\mathbf{f}(t')] = 6D\,\delta(t-t')$. We notice that this expression accounts for a delta correlation in time once a specific monomer is chosen, nevertheless, we must also account for some sort of null-correlation between the forces acting on different monomers, regardless of time. This is easily achieved by adding a Kronecker-delta such that the expression of the *fluctuation-dissipation theorem* becomes

$$\mathbb{E}[\mathbf{f}_n(t)\,\mathbf{f}_m(t')] = 6D\,\delta(t-t')\,\delta_{n,m}. \tag{1.31}$$

In this way, when considering two different monomers there is always no correlation between the random forces acting on them, whether they occur simultaneously or not; in fact $\mathbb{E}[\mathbf{f}_n(t)\,\mathbf{f}_m(t')] = 0$ whether $t = t'$ or $t \neq t'$, as long as $n \neq m$.

We are ready to find the equation of motion for the centre of mass of the polymer. By differentiating Eq. (1.28) we obtain: $d\mathbf{R}_{CM}(t)/dt = \frac{1}{N}\sum_{i=1}^{N} d\mathbf{R}_i(t)/dt$, which yields to:

$$\frac{d\mathbf{R}_{CM}(t)}{dt} = \frac{1}{N}\left\{-k(\mathbf{R}_1 - \mathbf{R}_2) - k\sum_{n=2}^{N-1}[(\mathbf{R}_n - \mathbf{R}_{n+1}) + (\mathbf{R}_n - \mathbf{R}_{n-1})] - k(\mathbf{R}_N - \mathbf{R}_{N-1}) + \sum_{n=1}^{N}\mathbf{f}_n\right\}, \tag{1.32}$$

---

[4] It can be shown (see [1]) that $k = \frac{3k_BT}{\gamma\, b^2}$ where $b = \mathbb{E}[\mathbf{r}^2]$ with $\mathbf{r}$ being the bond vector of two consecutive monomers in the chain. For our purposes we call k the generic spring constant to avoid making the notation cumbersome.

where we used Eqs. (1.29) and (1.30). It is easy to notice that the previous expression can be put in the following form:

$$\frac{d\mathbf{R}_{CM}(t)}{dt} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{f}_n,$$

which enables us to find a simple expression for $\mathbf{R}_{CM}(t)$, by integrating over the time interval $[0, t]$:

$$\mathbf{R}_{CM}(t) = \mathbf{R}_{CM}(0) + \frac{1}{N}\int_0^t dt' \sum_{n=1}^{N}\mathbf{f}_n. \tag{1.33}$$

The *diffusion coefficient* of a Rouse model polymer eventually emerges from the variance of the process $\mathbf{R}_{CM}(t)$, which reads:

$$\begin{aligned}
\mathrm{VAR}\left[\mathbf{R}_{CM}(t)\right] &\equiv \mathbb{E}\left[(\mathbf{R}_{CM}(t) - \mathbf{R}_{CM}(0))^2\right] \\
&= \int_0^t dt' \int_0^t dt'' \, \frac{1}{N^2} \sum_{m,n=1}^{N} \mathbb{E}\left[\mathbf{f}_n(t') \, \mathbf{f}_m(t')\right] \\
&= 6D \int_0^t dt' \int_0^t dt'' \, \frac{1}{N^2} \sum_{m,n=1}^{N} \delta_{m,n} \, \delta(t' - t'') \\
&= \frac{6D}{N} \int_0^t dt' \\
&= 6\left(\frac{D}{N}\right) t,
\end{aligned} \tag{1.34}$$

where we used the *fluctuation-dissipation theorem* (1.31). We therefore showed that, under the hypothesis made, the diffusion coefficient of the centre of mass of a polymer described by the Rouse model is

$$D_{CM} = \frac{D}{N} \tag{1.35}$$

where D is the diffusion coefficient of one monomer and N the size of the polymer. The expression obtained is physically reasonable: as N increases D_{CM} tends to 0 (a bigger polymer diffuses less than a smaller one) and when $N = 1$, $D_{CM} = D$ (the polymer becomes a monomer).

Chapter 2

Non-Gaussian effects in Brownian motion

In the previous section, we discussed Brownian motion through its *Langevin equations*, which allowed us to determine the diffusion coefficient of a Rouse model polymer in an aqueous medium. Here, instead, we want to investigate the possibility of a Brownian non-Gaussian diffusion, thus a displacement process whose variance depends linearly on time, yet the probability density function is non-Gaussian. This behaviour was discovered in a variety of experimental conditions and molecular dynamics simulations, consistently occurring due to the presence of an additional source of randomness beyond the fluctuating force. Although showing in a multitude of different contexts, we will focus on the Brownian non-Gaussian behaviour of a polymer in a diluted solution that can freely exchange monomers with a chemostat. In this case, the second source of randomness is provided by the fluctuating diffusion coefficient of the polymer centre of mass, leading to its position exhibiting Brownian non-Gaussian diffusion. The *Langevin equations* for Brownian motion, together with the Rouse model of the previous chapter, provide the natural environment for developing a theory of non-Gaussianity, allowing us to quantify deviations from normal behaviour.

2.1 The gran canonical polymer model

Let us now focus on the dynamics of polymers. We showed that the centre of mass of a polymer diffuses with the following diffusion coefficient $D_{CM} = D/N$ where D is the diffusion coefficient of each monomer and N the polymer size. Assuming the centre of mass of the polymer behaves like a monomer with diffusion coefficient D_{CM}, we are under standard conditions when N is fixed, and thus the diffusion is normal as emerges from Eq. (1.26). When the polymer is instead put into a chemical bath in the presence of other monomers, it randomly increases and decreases in size. In this situation, the variable N becomes a time-dependent random variable, imparting probabilistic characteristics to D_{CM}. The easiest way to model such a process is by considering a *birth-death process*, where the number $N(t)$ increases by one with a certain rate λ (the birth rate) and decreases by one with another rate μ (the death rate). We define the probability of changing N from the value i to the value j as $P_{i,j}(t) = P\{N(t+s) = j \,|\, N(s) = i\}$ with t being the time increment during which the variable updates its value. For an infinitesimal increment $dt > 0$, $P_{i,j}(dt)$ is assumed to satisfy the following properties (at least at the lowest order):

- $P_{i,i+1}(dt) = \lambda\, dt$ for $i \geq 0$ (a birth);
- $P_{i,i-1}(dt) = \mu\, dt$ for $i \geq 1$ (a death);
- $P_{i,i}(dt) = 1 - (\lambda + \mu)\, dt$ for $i \geq 0$ for $i \geq 1$ (no change).

Furthermore, in this model it is possible to approximate the autocorrelation time τ of $N(t)$ - which is the time it takes the random variable to lose memory of its previous value - as (see [7])

$$\tau = \frac{1+p}{(1-p)^2\mu}, \tag{2.1}$$

where $p \equiv \lambda/\mu$. We notice that as $p \to 1^-$, τ diverges and the polymer is said to be critic. This is the particular situation in which, as we shall see later, the polymer diffuses in a non-Gaussian way at all time t.

2.1.1 The subordinator process

After understanding how the fluctuating diffusion coefficient emerges from the dynamics of polymers in a grand canonical ensemble, we are ready to quantify non-Gaussian processes. Here we show a general mathematical approach for addressing these issues, known as *subordination*, and at the end of the section, we will apply these results to the grand canonical polymer model. The starting point is Eq. (1.22) in which we need to substitute the diffusive constant D with the stochastic process $D(t)$ (we omit the label CM to keep the notation simple), so that it becomes

$$X(t+dt) = X(t) + \sqrt{2D(t)}\, dW(dt) \quad \Rightarrow \quad dX(t) = \mathbb{N}\left(0, 2D(t)dt\right) \tag{2.2}$$

since we had defined $dW(dt) \equiv \mathbb{N}(0, dt)$ and used the identity $\alpha + \beta\mathbb{N}(m, \sigma^2) = \mathbb{N}(\alpha + \beta m, \beta^2\sigma^2)$ valid for any two numbers α and β [4]. At this point, we can make any assumption about the distribution of the random variable D. For simplicity we assign a steady-state distribution p_D^*, which could either be a probability density function or a probability mass function depending on whether $D(t)$ varies continuously or discretely. A steady-state distribution is one in which the moments remain constant over time (though the values the random variable take obviously change), so we can denote the mean and variance of $D(t)$ as $\mathbb{E}[D] \equiv \mu_D^*$ and $\mathrm{VAR}[D] \equiv (\sigma_D^*)^2$, respectively. In the context of the probabilistic mechanism of *subordination*, we introduce the *subordinator process* $S(t)$ defined as:

$$S(t) \equiv 2\int_0^t dt'\, D(t') \quad \Rightarrow \quad dS = 2D(t)dt, \tag{2.3}$$

and we call $X(t)$ the *subordinated process*. This tool gives a convenient parameterization of time such that we can express Eq. (2.2) in the following form

$$dX(t) = dW(dS) = \mathbb{N}(0, dS). \tag{2.4}$$

With this parameterization, we can switch from the deterministic variable t to a stochastic definition of time, since when a realization of $D(t)$ is made at time t, $S(t)$ will take on a different value according to its stochastic evolution.

Our goal is to determine the probability distribution function for the process $X(t)$, or at least, ascertaining whether it is Gaussian or not. In order to do so we introduce the *subordination formula* defined as

$$p_X(x, t|x_0) = \int_0^{+\infty} ds\, G_{BG}(x, s|x_0)\, p_S(s, t), \tag{2.5}$$

where $p_S(s, t)$ is the probability for s at time t, and $G_{BG}(x, s|x_0)$ is the Green's function for the Brownian-Gaussian (BG) diffusion associated in our cases with natural boundary conditions. According to Eq. (1.26) the Green's function $G_{BG}(x, s|x_0)$ is normal:

$$G_{BG}(x, s|x_0) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{(x-x_0)^2}{2s}}. \tag{2.6}$$

Let us now introduce the so-called *excess kurtosis* for the process $X(t)$ which is defined as the *kurtosis* minus 3:

$$k_X(t) - 3 \equiv \frac{\mathbb{E}\left[(X(t)-x_0)^4\right]}{\left(\mathbb{E}\left[(X(t)-x_0)^2\right]\right)^2} - 3. \tag{2.7}$$

This statistical tool serves as a Gaussian index:

- if $k_X(t) - 3 > 0 \Rightarrow$ the process $X(t)$ is *leptokurtic*: it has "fatter tails";
- if $k_X(t) - 3 = 0 \Rightarrow$ the process $X(t)$ is Gaussian[1];
- if $k_X(t) - 3 < 0 \Rightarrow$ the process $X(t)$ is *platykurtic*: it has "thinner tails".

To characterise its possible non-Gaussianity, we can thus compute the excess kurtosis for the process $X(t)$. First of all, let us manipulate, via a change of variable, Eq. (2.5) to find an expression for the m-th moment of the probability distribution of $X(t)$,

$$\begin{aligned}
\mathbb{E}\left[(X(t)-x_0)^m\right] &= \int_{-\infty}^{+\infty} dx\, p_X(x,t|x_0)\,(x-x_0)^m \\
&= \int_0^{+\infty} ds\, p_S(s,t) \int_{-\infty}^{+\infty} dx\, \frac{1}{\sqrt{2\pi s}} e^{-\frac{(x-x_0)^2}{2s}}\,(x-x_0)^m \\
&= \left(x' = \frac{x-x_0}{\sqrt{s}} \Rightarrow x = x'\sqrt{s} + x_0 \Rightarrow dx = dx'\sqrt{s}\right) \\
&= \int_0^{+\infty} ds\, p_S(s,t) \int_{-\infty}^{+\infty} dx'\, \frac{e^{-\frac{(x')^2}{2}}}{\sqrt{2\pi s}} \sqrt{s}\, s^{\frac{m}{2}}\,(x')^m \\
&= \int_0^{+\infty} ds\, p_S(s,t)\, s^{\frac{m}{2}} \int_{-\infty}^{+\infty} dx'\, \frac{e^{-\frac{(x')^2}{2}}}{\sqrt{2\pi}}\,(x')^m \\
&= G_{BG}^{(m)} \int_0^{+\infty} ds\, p_S(s,t)\, s^{\frac{m}{2}} \\
&= G_{BG}^{(m)}\, \mathbb{E}[S(t)^{\frac{m}{2}}],
\end{aligned}$$

where $G_{BG}^{(m)}$ is the m-th moment of a unit normal distribution; therefore, we consider its value known for any m. Hence, we found a way to express the moments of the distribution of $X(t)$ through the moments of the distribution of $S(t)$. It is now easy to rewrite the excess kurtosis for the process $X(t)$ as follows

$$\begin{aligned}
k_X(t) - 3 &= \frac{G_{BG}^{(4)}\,\mathbb{E}\left[S(t)^2\right]}{\left(G_{BG}^{(2)}\right)^2 \left(\mathbb{E}[S(t)]\right)^2} - 3 \\
&= 3\frac{\mathbb{E}\left[S(t)^2\right] - \left(\mathbb{E}[S(t)]\right)^2}{\left(\mathbb{E}[S(t)]\right)^2} \\
&= 3\frac{\sigma_S^2(t)}{\mu_S^2(t)}
\end{aligned}$$

---

[1] In fact, $\mathbb{E}\left[(X(t)-x_0)^4\right] = 3\sigma^2$ for a Gaussian process, so that: $k_X(t) - 3 = 3 - 3 = 0$.

2.1.2 Excess kurtosis of the subordinator process

In order to express the excess kurtosis for $X(t)$ we compute its mean and variance. By definition:

$$\begin{aligned}\mu_S(t) &= \mathbb{E}\left[2\int_0^t dt'\, D(t')\right]\\ &= 2\int_0^t dt'\, \mathbb{E}[D(t')]\\ &= 2\mu_D^* t,\end{aligned} \tag{2.8}$$

note that we assumed $D(t)$ follows a steady-state distribution. Now, in order to find the variance of $S(t)$ let us first calculate the quantity $\mathbb{E}[S(t)^2]$:

$$\begin{aligned}\mathbb{E}[S(t)^2] &= \mathbb{E}\left[4\int_0^t dt'' \int_0^t dt'\, D(t'')\, D(t')\right]\\ &= 4\int_0^t dt'' \int_0^t dt'\, \mathbb{E}[D(t'')\, D(t')]\\ &= 2\left\{4\int_0^t dt'' \int_0^{t'=t''} dt'\, \mathbb{E}[D(t'')\, D(t')]\right\}\end{aligned} \tag{2.9}$$

where, instead of integrating over a square of side length t, we summed twice the integral over the upper triangle region formed by the bisector of the first and third quadrants (the line $t' = t''$), since the two triangles are equal in area by construction (we selected the upper triangle to ensure that $t'' > t'$). Let us find an expression for $\mathbb{E}[D(t'')\, D(t')]$. Provided that τ is the autocorrelation constant of the process $D(t)$, we expect that for $t' = t''$ it becomes $\mathbb{E}[D(t'')\, D(t')] = \mathbb{E}[D^2]$, conversely, for $t'' - t' >> \tau$ the quantity $D(t'')$ loses memory of $D(t')$ so that $D(t'')$ and $D(t')$ are temporally uncorrelated. The latter argument can be formalised as follows: $\mathbb{E}[D(t'')\, D(t')] = (\mathbb{E}[D])^2$, given that $\mathrm{COV}[D(t''),\, D(t')] = 0$ for $t'' - t' >> \tau$. Finally, we propose that the transition from the first situation to the second occurs via exponential decay. The following expression satisfies all our requirements:

$$\begin{aligned}\mathbb{E}[D(t'')\, D(t')] &= \left\{\mathbb{E}[D^2] - (\mu_D^*)^2\right\} e^{-2\frac{t''-t'}{\tau}} + (\mu_D^*)^2\\ &= (\sigma_D^*)^2\, e^{-2\frac{t''-t'}{\tau}} + (\mu_D^*)^2\,.\end{aligned} \tag{2.10}$$

We can now substitute Eq. (2.10) in Eq. (2.9) and evaluate the integral

$$\begin{aligned}\mathbb{E}[S(t)^2] &= 8\int_0^t dt'' \int_0^{t''} dt' \left\{(\sigma_D^*)^2\, e^{-2\frac{t''-t'}{\tau}} + (\mu_D^*)^2\right\}\\ &= 8\,(\sigma_D^*)^2 \int_0^t dt''\, e^{-\frac{2t''}{\tau}}\, \frac{\tau}{2}\left[e^{2\frac{t'}{\tau}}\right]_0^{t''} + 4\,(\mu_D^*)^2\, t^2\\ &= 4\,(\sigma_D^*)^2\, \tau \int_0^t dt'' \left(1 - e^{-2\frac{t''}{\tau}}\right) + 4\,(\mu_D^*)^2\, t^2\\ &= 4\,(\sigma_D^*)^2\, \tau \left[t'' + \frac{\tau}{2}\left(e^{-2\frac{t''}{\tau}}\right)\right]_0^t + 4\,(\mu_D^*)^2\, t^2\\ &= 4\,(\sigma_D^*)^2\, \tau t + 2\,(\sigma_D^*)^2\, \tau^2 \left(e^{-2\frac{t}{\tau}} - 1\right) + 4\,(\mu_D^*)^2\, t^2.\end{aligned}$$

Consequently, the variance of $S(t)$ is:

$$\mathrm{VAR}[S(t)] = \mathbb{E}[S(t)^2] - \mathbb{E}[S(t)]^2 = 4\,(\sigma_D^*)^2\, \tau t + 2\,(\sigma_D^*)^2\, \tau^2 \left(e^{-2\frac{t}{\tau}} - 1\right), \tag{2.11}$$

thus, the excess kurtosis can be expressed as

$$k_X - 3 = 3\frac{2\tau\left(\sigma_D^*\right)^2\left[2t+\tau\left(e^{-2\frac{t}{\tau}}-1\right)\right]}{4\left(\mu_D^*\right)^2 t^2}. \tag{2.12}$$

From the plot of the excess kurtosis (Fig. (2.1)) we find out the behaviour of the process $X(t)$. We see that as $t''-t' << \tau$, the excess kurtosis tends to a positive constant[2], conversely for $t''-t' >> \tau$, $k_X(t)-3 \to 0$. According to the behaviour of the excess kurtosis, we observe that for $t''-t' << \tau$ the process $X(t)$ exhibits leptokurtic behaviour and is thus non-Gaussian, while as time passes the non-Gaussian feature diminish and the process quickly converges to standard Brownian motion. Moreover, we can now explain the observation made at the end of subsection (2.1): from Eq. (2.1) we see that if $p \to 1^-$, $\tau \to +\infty$ (autocorrelation time of $N(t)$). This implies that if the polymer is critic, its motion remains constantly leptokurtic, since $k_X(t)-3 \to +\infty$ for all $t>0$ [7].

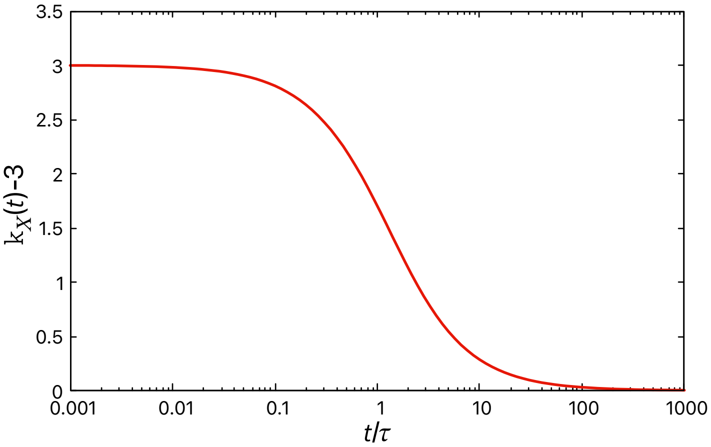

Figure 2.1: A plot of the excess kurtosis for the process $X(t)$ for $\mu_D^*/\sigma_D^*=1$.

2.1.3 The Jensen's inequality

We want now to give some more details about the shape of a leptokurtic distribution for a subordinated process. To achieve this, we can use the Jensen's inequality to show that the "tail effect" (responsable for faster diffusion) is accompanied by an excess of probability in the centre (responsable for slower diffusion). Clearly, these two possibilities for $X(t)$ are both triggered by the choice of the value of $D(t)$ taken from the ensamble of the possible ones, regulated by the steady-state distribution p_D^*. The Jensens's inequality says that for a real-valued μ-measurable function f on a sample space Ω and a convex function ϕ on the real numbers we have:

$$\int_\Omega d\mu\,(\phi\circ f) \geq \phi\left(\int_\Omega d\mu f\right). \tag{2.13}$$

The inequality becomes strict if ϕ is strictly convex and the measure μ is not induced by a constant random variable. By making the following substitutions, we can derive the necessary conclusions from the inequality:

- $\phi(s) \equiv G_{BG}(x_0, s|x_0) = \frac{1}{\sqrt{2\pi s}}$ which is a strict convex function of s;
- $d\mu \equiv ds\, p_S(s,t)$;
- $f(s) \equiv s$.

---

[2] It is easy to find that this constant is equal to $3\frac{(\sigma_D^*)^2}{(\mu_D^*)^2}$, just by taking the limit $\to 0$ of Eq. (2.12).

With these specifications, we obtain:

$$\begin{aligned} p_X(x_0,t|x_0) &= \int_0^{+\infty} ds\, G_{BG}(x_0,s|x_0)p_S(s,t) \\ &> G_{BG}\left(x_0,\, \int_0^{+\infty} ds\, p_S(s,t)s|x_0\right) \\ &= G_{BG}\left(x_0,\, \mathbb{E}[S(t)]|x_0\right), \end{aligned}$$

for all time t. The inequality was used at the point x_0, but since both $p_X(x,t|x_0)$ and $G_{BG}(x,s|x_0)$ are continuous functions around that point, there must exist a neighbourhood I of the centre x_0 in which the inequality remains valid:

$$p_X(x,t|x_0) > G_{BG}\left(x,\, \mathbb{E}[S(t)]|x_0\right) \quad \text{for all } x \in I \text{ and all time } t. \tag{2.14}$$

Therefore, not only there is a tail effect but also an excess of probability in the centre. The profile of the distribution of a non-Gaussian Brownian motion is given in Fig. (2.2) compared to the standard Gaussian one.

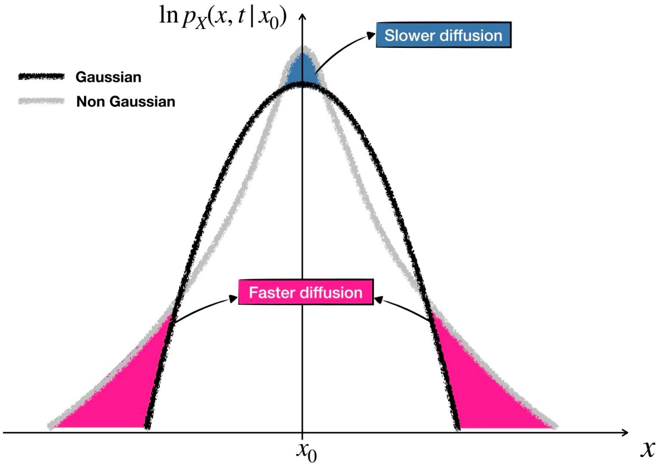

Figure 2.2: Comparison between Gaussian and non-Gaussian probability density functions for subordination processes. The two distributions share the same mean and variance (from [7]).

Chapter 3

Polymer translocation

The ability of polymer molecules to undergo large conformational changes, without loosing their topological connectivity, makes them the fundamental biological structure of life. Polymers directly regulate the most important macromolecular processes by acting in a number of crowded environments such as cells. So far, we have discussed the way polymers diffuse through both a sterile medium and one containing other monomers. To explore other biological processes where polymers are central, we will now focus on *polymer translocation*, which refers to the passage of polymer chains across a small channel in a membrane [6]. Many different phenomena exploit this ability of polymers; some of them are the passage of the filament of mRNA (messenger RNA) through nuclear pore complexes to carry on protein synthesis, the injection of DNA from a virus into a host cell, and also gene swapping through pili. Furthermore, scientists use this effect to decipher the sequence of DNA using current measurements through the translocation channel. Although polymer translocation emerges in different contexts and is influenced by the specific chemical details of both the molecule and the environment, our discussion here will focus on its universal physical features. In this chapter, starting from thermodynamic considerations, we will find a way to model the process of translocation using the mathematics previously developed.

3.1 Generalities of translocation

Let us first provide more biological details of this phenomenon. Translocation naturally occurs in cellular environments because the membranes of both eukaryotic and prokaryotic cells are abundant in pore proteins. In fact, one of the first pores to be implemented in an experimental setup was the pore-forming protein α-hemolysin. This specific pore has a width of 1.4 nm at its narrowest point, allowing only the passage of single-stranded DNA or RNA, as well as unfolded protein chains [6]. Generally speaking, biological pores rarely exceed 2 nm in diameter; moreover, they are susceptible to external parameters such as temperature and pH, making them disadvantageous from an experimental point of view. This highlights the need to develop synthetic pores that can be precisely engineered for adaptability to various situations. Regardless of the specific pore used, the polymer translocation process can be divided into three main steps: *drift-diffusion*, *capture*, and *translocation*.

1. In the *drift-diffusion* step, the polymer diffuses not only due to random collisions of the solvent molecules (exhibiting the Brownian motion of the polymer centre of mass) but also for the presence of a possible external force field (drift). The polymer behaviour in this step therefore depends both on the polymer structure and the background fluid.

2. The *capture* process refers to the embedding of the polymer into the pore mouth. This step is regulated by many contributing factors of both a chemical and a physical nature. However, their net effect generates a sucking force at the pore entrance, which facilitates translocation.

3. Finally, the *translocation* process itself refers to the transit of the polymer through the hole, after one segment has already reached the receiver compartment.

In this thesis, we will concentrate solely on the final stage. We will examine it from both energetic and kinematic perspectives to derive key insights and results.

3.2 Thermodynamic aspects of translocation

As we just discussed, before a polymer starts to translocate through a pore it first needs to be threaded inside. However, we will disregard the capture process and consider the chain already embedded in the membrane. We start by labelling the monomers from 1 to N, with monomer 1 being the first to enter the pore. In this way, we measure the progress of the chain through the pore, by means of the coordinate m which represents the number of monomers that have already translocated into the receiver compartment (trans); consequently, $N - m$ monomers are in the donor compartment (cis). The coordinate m is a stochastic variable, as it represents the diffusion of the monomer across the pore immersed in a solvent (whose composition will be specified later). We therefore expect m to follow the Langevin equations of motion presented in subsection (1.2.1); however, in this case, there is not only the effect of the solvent molecules randomly colliding on the particle, but also an *entropic force* effect. Besides from possible energetic contributions, part of this deterministic force arises from the difference in entropy between the receiver and donor compartments; indeed, a greater number of microscopic configurations for one of the two polymer tails provides greater entropy with respect to the other. For this reason, the coordinate m experiences a sucking deterministic force that aims to maximise the entropy. In order to handle this force, let us focus on the thermodynamic aspects of the process. For being immersed in a thermal reservoir at absolute temperature T, we will consider the free energy F of this quasi one-dimensional diffusion problem given by

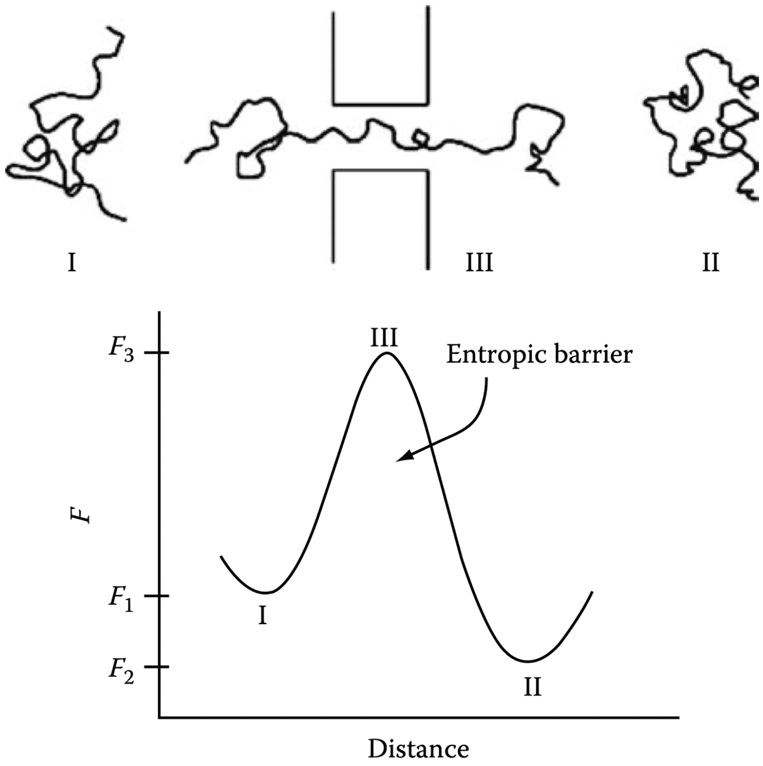

Figure 3.1: Expected profile of the free energy for polymer translocation (from [5]).

$$F(T, m, N) \equiv U(S(T, m, N), m, N) - TS(T, m, N).$$

This expression is known as the fundamental relation [2] in the free energy representation since it contains all of the equilibrium thermodynamic properties of the system. In this way as the entropy increases, we expect a decrease in the free energy, if U remains constant. Now, by differentiating the latter definition, the fundamental equation of the system in the free energy representation reads

$$dF = -SdT - f dm + \mu dN. \tag{3.1}$$

From this expression we obtain the entropic force acting on the monomer along the pore, which is defined as

$$f(T, m, N) \equiv - \left(\frac{\partial F(T, m, N)}{\partial m} \right)_{T,N}. \tag{3.2}$$

This force must be added to the Langevin equation of motion to account for the discussed entropic effect; this will allow us to evaluate the probability density function of the stochastic variable $m(t)$. Hence, our goal is to first compute the fundamental relation of the tethered polymer to derive the entropic force, and then attempt to solve its equation of motion.

Before finding the exact formula for $F(T, m, N)$ we can depict its expected profile in Fig. (3.1). We see that, as the polymer begins to translocate, the free energy increases since there are fewer microscopical configurations available (thus a lower entropy) with respect to the free molecule. We shall call $(F_3 - F_1)$ the *entropic barrier* [5] (altought it is a free energy barrier) which acts against polymer translocation. In fact, by definition (3.2) the entropic force points in the direction of maximum free energy decrease; thus the polymer must negotiate with the entropic barrier in order for it to successfully arrive at the opposite side of the pore. At the end of this chapter we will indeed evaluate the *translocation time* (under specific conditions) which is the average first passage time for the stochastic translocation process [5].

3.3 Free energy of a polymer

We can calculate the free energy of a polymer molecule using

$$F = -k_B T \ln Z \tag{3.3}$$

where Z is the partition sum of the system, i.e. the number of microscopic conformations it can take; and k_B is the Boltzmann constant [3]. In this section, we first aim to write an expression for the partition sum of a polymer, by considering a *random walk* model and a *self-avoiding walk* model, next to compute the free energy of the system. We will discuss both the cases of a free polymer and a tethered one (e.g., to a biological membrane).

3.3.1 The random walk model for free polymers

We discussed the *random walk* model in section (1.1), in relation to the motion of a colloidal particle dispersed in a medium. In this case, to account for polymer conformations, the random walk can be thought of as snapshots of the polymer conformation taken at fixed time. Although it is given a different interpretation, the mathematics of the model remains unchanged. Indeed, we assume that the links of the polymer chain do not interact and can freely overlap with each other, so all configurations share the same energy. As a result, we expect Z to depend solely on the spatial dimention of the random walk and the number of steps N:

$$Z_{RW}(N) = \overline{\mu}^N,$$

where $\overline{\mu} \equiv 2d$ is called the *connective constant* with d being the dimension of the random walk [8]. For instance, if we think of a free polymer in one dimension, the partition sum reads $Z(N) = 2^N$, reflecting the fact that there are two possible moves at each step. Generally speaking, the connective constant is the *coordination number* of the lattice model considered for the random walk, i.e. the number of available steps immediately surrounding any position. The free energy associated to a free polymer molecule in a bath at absolute temperature T modelled as a random walk is thus

$$F_{RW}(T, N) = -N k_B T \ln \overline{\mu}.$$

Now, from Eq. (3.1) we see that the chemical potential can be defined as:

$$\begin{aligned}\mu(T) &\equiv \left(\frac{\partial F_{RW}(T,N)}{\partial N}\right)_T \\ &= \frac{F_{RW}(T,N)}{N} \\ &= -k_B T \ln \overline{\mu},\end{aligned}$$

it represents the work required to add a step to the random walk. The expression of the chemical potential yields the following identification:

$$\overline{\mu} = e^{-\mu/k_B T}, \tag{3.4}$$

therefore the connective constant corresponds to the inverse of the *fugacity* of the system.

3.3.2 The self-avoiding walk model for free polymers

A more realistic lattice model for linear polymers configurations takes into account the self-avoidance of monomers due to steric interactions, and it is called the *self-avoiding walk* (SAW) model. In this case, the links of the polymer chain interact and do not overlap in space. This model, introduces a power law to the number of conformations of the polymer:

$$Z_{SAW}(T,N) = \overline{\mu}^N \, N^{\gamma'-1}, \tag{3.5}$$

where γ' is a critical exponent that depends on the nature of the polymer and the background fluid, and also on any of its spatial restriction [5]. The associated free energy is thus

$$F_{SAW}(T,N) = -Nk_B T \ln \overline{\mu} - k_B T(\gamma' - 1)\ln N;$$

and the chemical potential becomes:

$$\begin{aligned}\mu(T) &= -k_B T \ln \overline{\mu} - k_B T \frac{\gamma' - 1}{N} \\ &= (N >> 1) \\ &\approx -k_B T \ln \overline{\mu}.\end{aligned}$$

We see that the identification (3.4) remains valid also for this model; therefore the free energy associated to a free polymer molecule in a bath at absolute temperature T modelled as a self-avoiding walk is

$$\begin{aligned}F_{SAW}(T,N) &= -Nk_B T \ln e^{-\mu/k_B T} - k_B T(\gamma' - 1)\ln N \\ &= N\mu + k_B T(1 - \gamma')\ln N,\end{aligned}$$

where $\gamma' \approx 1.16$ in good solutions. We will specify in a moment the meaning of "good solution", but first let us analyse the introduction of the power law $N^{\gamma'-1}$ to the partition sum. It is easy to see that the greater the exponent, the greater the number of possible configurations of the polymer. Thus, if the polymer is spatially restricted, we expect the critical exponent to decrease to account for a lower number of microscopic conformations.

3.3.3 The self-avoiding walk model for tethered polymers

We are now ready to consider a polymer attached to a flat wall, which is the scenario of the two polymer tails in the translocation process. In this case, we expect a decrease in the number of allowed configurations due to the prohibition of some polymer conformations, as shown in Fig. (3.2). To address this problem, we focus on a polymer rooted to a wall, representing half of the translocating

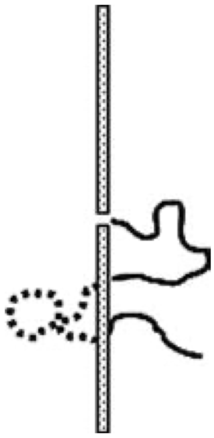

Figure 3.2: Prohibited polymer conformation due to the impenetrability of the surface of the membrane. (from [5]).

polymer, and we introduce the critical exponent γ''. The partition sum and the free energy, in the case of a *self-avoiding walk* model for the tethered polymer, thus become:

$$Z_{half}(N) = e^{-N\mu/K_BT}\, N^{\gamma''-1} \quad \Rightarrow \quad F_{half}(T,N) = N\mu + k_BT(1-\gamma'')\ln N, \tag{3.6}$$

where we dropped the "SAW" label for an easier notation. The effective values of γ'' for the two major experimental cases are:

- $\gamma'' \approx 0.69$, good solutions (high salt);
- $\gamma'' = 1$, bad solutions (low salt).

What we physically mean by a "good solution" is a solvent able to surround each monomer of the chain by forming chemical bonds. In this way, the polymer adopts linear conformations, preventing it from overlapping with itself. Conversely, when a polymer is dispersed in a low salt solution or "bad solution", it takes globular conformations, which means it tends to form a compact "ball" of monomers. Moreover, as said earlier, the critical exponent depends not only on the background fluid but also on the spatial restriction of the polymer. In fact, a free polymer has a higher exponent than a rooted one, assuming they are in the same solution (see the case of $\gamma' \approx 1.16$ for a free polymer versus $\gamma'' \approx 0.69$ for a rooted one, provided it is dispersed in a "good solution"). Again, the former situation can take more conformations than the latter. In this chapter, we will always consider the self-avoiding walk model for polymers with a "good solution" critical exponent.

3.4 Free energy of a translocating polymer

In this section, we aim to evaluate the free energy of a translocating polymer across a membrane. The theoretical approach to address this problem was suggested by Sung and Park in 1996 [6]. They deduced the free energy of the chain by summing the free energies of the two polymer tails of sizes $N-m$ and m belonging to the donor and receiver regions, respectively. Indeed, instead of thinking of one polymer, we can model the problem with two half-chains embedded to a flat wall. Thus, we can safely assume that the total number of microstates factorises with respect to those of the donor and receiver compartments:

$$Z(m,N) = Z_{half}(N-m)\, Z_{half}(m),$$

where Z_{half} refers to the first Equation in (3.6) which contains the critical exponent for a rooted polymer. Using Equations (3.3) and the second in (3.6), the total number of conformations reads:

$$\begin{aligned} F(T,m,N) &= -k_BT\ln Z(m,N) \\ &= -k_BT\ln Z_{half}(N-m) - k_BT\ln Z_{half}(m) \\ &= F_{\text{half}}(T,N-m) + F_{half}(T,m) \\ &= (N-m)\mu_{cis} + k_BT\left(1-\gamma''_{cis}\right)\ln(N-m) + m\mu_{trans} + k_BT\left(1-\gamma''_{trans}\right)\ln m \\ &= k_BT\left(1-\gamma''_{trans}\right)\ln m + k_BT\left(1-\gamma''_{cis}\right)\ln(N-m) - m\Delta\mu + N\mu_{cis}, \end{aligned}$$

where γ''_{cis} and γ''_{trans} are the critical exponents of the donor and receiver compartments, respectively; and $\Delta\mu \equiv \mu_{cis} - \mu_{trans}$ is the difference in the chemical potential per polymer segment across the membrane, which we call the *driving force*. For simplicity, we consider two identical good solutions in the two regions, thus $\gamma''_{cis} = \gamma''_{trans} = \gamma'' \approx 0.69$; and disregard the unnecessary constant term $N\mu_{cis}$. Therefore, the free energy of the translocating polymer becomes the following

$$F(T,m,N) = k_BT\left(1-\gamma''\right)\ln[m(N-m)] - m\Delta\mu. \tag{3.7}$$

A plot of this expression (divided by k_BT) against the variable m is given in Fig. (3.3) for a polymer of length $N = 100$. We immediately notice that the graph qualitatively resembles Fig. (3.1), displaying an *entropic barrier* that the polymer needs to overcome. Clearly the process is completed as $m = N = 100$, by construction of the coordinate m. The plot shows three different scenarios given three values for the quantity $\Delta\mu/k_BT$, that arises for many reasons, such as the difference in the concentration of the polymer, the pH gradients, or the complexation with other macromolecules [5]. The green line represents the conditions for a potentially successful translocation process: the entropic barrier is already overcome when just a relatively small number of monomers m^* have translocated (from the plot we can estimate that $m^* \approx 20$); and starting from there, the free energy of the system decreases, therefore the process becomes spontaneous (since it tends to lower the free energy). The blue line represents the case of a null driving force, which reasonably gives a symmetric situation in which neither compartment is favored by the polymer. Finally, the red plot illustrates a potentially impossible translocation: even if the molecule manages to complete the process, the trans region is only metastable, so the polymer would tend to return to the donor compartment. In summary, a polymer translocation process can energetically occur only if the driving force $\Delta\mu$ is positive, meaning the chemical potential in the donor compartment is higher than in the receiver compartment.

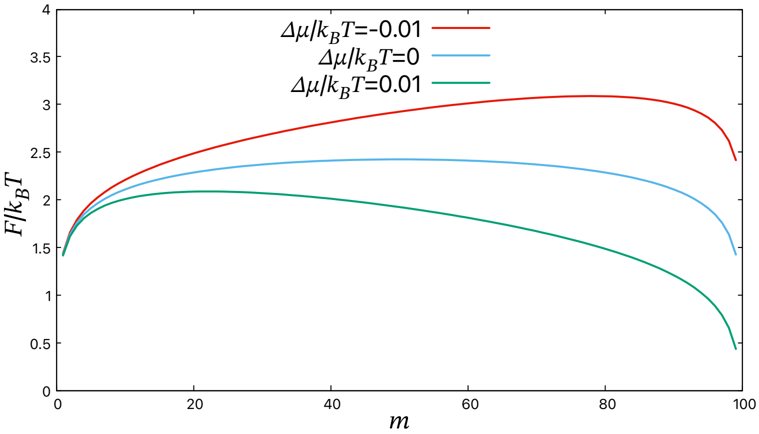

Figure 3.3: Plot of Eq. (3.7) against m in the case of positive (green line), null (blue line) and negative (red line) driving force.

The above discussion depicts the general behavior of a translocating polymer through a pore in which both the two terms in the expression of the free energy (3.7) are relevant. However, in most of the single-molecule electrophysiology experiments, the driving force is enormous with respect to the entropic barrier term [5] (the first appearing in Eq. (3.7)). For such strong driving forces, the free energy simply becomes

$$F(T, m, N) \simeq -m\Delta\mu. \tag{3.8}$$

The free energy landscape is thus a ramp with negative slope (if $\Delta\mu > 0$) indicating that translocation occurs without the need to overcome an entropic barrier, since the process is basically spontaneous for all m.

3.5 Translocation kinetics

Having an expression for the free energy of the system allows us to derive its thermodynamic properties. In fact, using this approach as stated at the beginning of this chapter, we can write a *Langevin equation* for the coordinate m of the translocating polymer for the study of the kinetics of the phenomenon. As previously said, the polymer molecule not only feels a random fluctuating force due to the collisions of the surrounding water molecules, but also a deterministic entropic force whose definition is given in (3.2). We will analyse the simple case of a dominant driving force over the entropic barrier term, whose behaviour is depicted in Eq. (3.8). In this way the entropic force simply reads:

$$f(T, m, N) = \Delta\mu.$$

Thus, by means of Eq. (1.27) written for the one-dimensional coordinate m, the *overdamped Langevin equation* for the process of polymer translocation (in the case of a dominant driving force) is:

$$\frac{dm}{dt} = \frac{\Delta\mu}{\gamma} + \mathrm{f}(t) \quad \text{or equivalently} \quad m(t + dt) = m(t) + \frac{\Delta\mu}{\gamma} dt + \sqrt{2D}\, dW(dt), \tag{3.9}$$

where $\mathrm{f}(t)$ is a zero-mean force that obeys to the *fluctuation-dissipation theorem*:

$$\mathbb{E}[\mathrm{f}(t)\, \mathrm{f}(t')] = 2D\, \delta(t - t').$$

Using the update form of Eq. (3.9) and the same arguments of subsection (1.2.1), we can infer that the process $m(t)$ is Gaussian, thus we only need its mean and variance to fully describe it.
We begin by integrate the differential *overdamped Langevin equation* over the time interval $[0, t]$, having $m(0) = m_0$:

$$m(t) = m_0 + \frac{\Delta\mu}{\gamma} t + \int_0^t dt'\, \mathrm{f}(t').$$

By averaging this result we find the mean of the process:

$$\mathbb{E}[m(t)] = m_0 + \frac{\Delta\mu}{\gamma} t.$$

Consequently, the calculation for the variance yields

$$\begin{aligned} \mathrm{VAR}[m(t)] &\equiv \mathbb{E}\left[(m(t) - \mathbb{E}[m(t)])^2\right] \\ &= \mathbb{E}\left[\left(\int_0^t dt'\, \mathrm{f}(t')\right)^2\right] \\ &= \int_0^t dt' \int_0^t dt''\, \mathbb{E}\left[\mathrm{f}(t')\, \mathrm{f}(t'')\right] \\ &= 2D\, t. \end{aligned}$$

We thus find

$$m(t) = \mathbb{N}\left(m_0 + \frac{\Delta\mu}{\gamma} t,\ 2D\, t\right),$$

which is similar to the solution of a Brownian motion but for having a linearly time-dependent mean. This behaviour well describes the phenomenon, since for a translocation process to be completed, the number of monomers in the receiver side (m) must reach the total number of monomers that make up the chain (N). A simulation sketch of the process is portrayed in Fig. (3.4). We can interpret the graph by picturing a very sharp Gaussian centred at m_0 that moves upward linearly over time, widening its variance, until it reaches N.

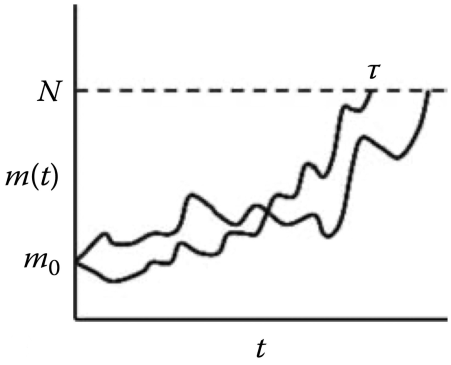

Figure 3.4: Illustration of two trajectories showing the time evolution of the number of monomers translocated into the receiver compartment (from [5]).

Finally, let us provide an expression for the translocation time, which is the first passage time τ. We ask for the time $t = \tau$ such that the mean of $m(t)$ reaches the value N, so that we are sure the translocation is successfully completed (clearly it is an approximation since $m(t) = N$ could be reached earlier or later). Therefore, by solving for τ the equation $\mathbb{E}\left[m(\tau)\right] = N$, we find

$$\tau = \gamma \frac{N - m_0}{\Delta\mu}.$$

By using the Einstein's relation: $\gamma D = k_B T$ (see, e.g, [4]), the translocation time becomes:

$$\tau = \frac{k_B T}{D} \frac{N - m_0}{\Delta\mu},$$

whose dependencies are physically meaningful.

Conclusions

In this thesis, we managed to provide a general outlook on polymers dynamics, by studying their diffusion through the crowded cell environment and their ability to translocate across a membrane. We obtained some important results. Starting from the general Langevin equation, we applied the Rouse model to describe polymers diffusion in a medium, yielding the expression $D_{CM} = D/N$ for the diffusion coefficient of the polymer centre of mass (being D that of a monomer and N the polymer size). We then implemented this result in the gran canonical polymer model to study its dynamics in a solvent that contains additional monomers. In that work, we demonstrated the non-Gaussian characteristics of the position probability density function, providing a broad perspective on diffusion processes.

In the last chapter, we addressed the problem of polymer translocation. First we introduced the concept of entropic force, next we wrote the overdamped Langevin equation for the coordinate m, showing its Gaussian behaviour.

A further step could be to investigate whether the Gaussian nature of the probability density function of m remains valid when the two regions contain monomers, rather than just solvent. At this point, it would be interesting to compare the translocation time with the one we identified for Gaussian translocation, to observe whether the process occurs faster or slower.

We believe that this work overall effectively summarises our understanding of polymer dynamics and offers valuable insights for examining stochastic processes in biological environments.

Bibliography

[1] W.J. Briels. *Theory of Polymer Dynamics.* Oct 1998. URL: `https://cbp.tnw.utwente.nl/PolymeerDictaat/polymerdynamics.pdf`.

[2] Herbert B. Callen. *Thermodynamics and an Introduction to Thermostatistics.* Wiley, 2nd edition, 1985.

[3] M. Falcioni, A. Vulpiani. *Meccanica Statistica Elementare – I fondamenti.* Springer-Verlag Italia, 2015.

[4] Daniel T. Gillespie. The mathematics of brownian motion and johnson noise. *American Journal of Physics*, (6):891–921, 1996. `doi:10.1119/1.18210`.

[5] M. Muthukumar. *Polymer translocation.* 2011. `doi:10.1201/b10901`.

[6] 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. `doi:10.1039/C4SM01819B`.

[7] Vittoria Sposini, Sankaran Nampoothiri, Aleksei Chechkin, Enzo Orlandini, Flavio Seno, and Fulvio Baldovin. Being heterogeneous is disadvantageous: Brownian non-gaussian searches. *Phys. Rev. E*, 109:034120, Mar 2024. URL: `https://link.aps.org/doi/10.1103/PhysRevE.109.034120`, `doi:10.1103/PhysRevE.109.034120`.

[8] Carlo Vanderzande. *Lattice Models of Polymers.* Cambridge Lecture Notes in Physics. Cambridge University Press, 1998.