Chromatin Structure and Dynamics : Insights from a Spectral Study of Critical Interacting Polymers

Timothy Földes

To cite this version:

Timothy Földes. Chromatin Structure and Dynamics : Insights from a Spectral Study of Critical Interacting Polymers. Condensed Matter [cond-mat]. Sorbonne Université, 2023. English. NNT : 2023SORUS582 . tel-04482302v2

HAL Id: tel-04482302
https://theses.hal.science/tel-04482302v2

Submitted on 28 Feb 2024

SORBONNE UNIVERSITÉ

école doctorale 564 Physique en Île-de-France (EDPIF)
Laboratoire de physique théorique de la matière consendée (LPTMC)

CHROMATIN STRUCTURE AND DYNAMICS: INSIGHTS FROM A SPECTRAL STUDY OF CRITICAL INTERACTING POLYMERS

Par TIMOTHY FÖLDES

Thèse de doctorat de PHYSIQUE

Dirigée par MARIA BARBI

Présentée et soutenue publiquement le 21/12/2023

Devant un jury composé de :

DANIEL JOST, DR	ENS Lyon	Rapporteur
JOHN MARKO, PR	Northwestern University	Rapporteur
JEAN-LOUIS BARRAT, PR	Université Grenoble Alpes	Examinateur
ANTOINE COULON, DR	Institut Curie	Examinateur
MARIE JARDAT, PR	Sorbonne Université	Examinatrice
JUDITH MINÉ-HATTAB, CR	Sorbonne Université	Membre invité
MARIA BARBI, PR	Sorbonne Université	Directrice de thèse

RESUMÉ

L'objectif principal de cette thèse est de réaliser une étude complète de modèles de polymères afin d'établir des résultats de référence qui puissent améliorer l'analyse des données expérimentales sur la chromatine. Cette recherche porte à la fois sur les aspects d'équilibre et de dynamique, abordant des questions fondamentales sur le comportement des polymères dans les systèmes biologiques.

Dans la première partie de la thèse, l'accent est mis sur les propriétés d'équilibre, fournissant un cadre théorique pour l'étude des données FISH séquentielles. Ceci inclut l'introduction de concepts essentiels de la physique des polymères à l'équilibre, tels que les chaînes gaussiennes. Il explore également le concept des modes de Rouse dans une nouvelle perspective, en le reliant aux techniques d'analyse du signal. Des simulations numériques de polymères auto-interagissant sont réalisées et analysées, ce qui conduit à la définition d'un paramètre d'ordre basé sur les spectres qui caractérise les états de repliement des polymères, jusqu'à l'établissement d'un diagramme de phase précis pour la transition de phase *coil-globule*.

L'approche spectrale est finalement appliquée aux données FISH séquentielles avec deux résultats distincts. Premièrement, elle permet de détecter une signature spectrale indiquant un comportement critique des polymères, et validant ainsi l'utilité des modèles simples de polymères dans l'analyse de la chromatine. Deuxièmement, grâce à la mise en place d'un protocole d'analyse de données spécifique, elle permet de reveler des boucles non identifiées auparavant dans les mêmes données, offrant ainsi de nouvelles perspectives sur la étude de l'organisation fine de la chromatine.

La deuxième partie de la thèse s'intéresse aux aspects dynamiques de la transition de phase coil-globule. Cette partie commence par introduire les concepts relatifs à la dynamique stochastique et à la dynamique des polymères, y compris l'équation de Langevin et le modèle (dynamique) de Rouse. Des simulations extensives de dynamique de Langevin de polymères auto-interagissants sont présentées, ainsi qu'un *modèle de Rouse interagissant* qui tient compte de la transition de phase coil-globule et des effets de volume exclu. L'introduction de contraintes topologiques dans les simulations met en évidence la robustesse de la dynamique de Rouse, qui n'est altérée de manière significative que dans les phases de globules denses.

Dans le dernier chapitre, les nouveaux acquis sont utilisés pour expliquer des données apparemment paradoxales issues d'une étude récente de la dynamique de la chromatine de *Drosophile*, et donnent à nouveau des indications sur la pertinence d'un modèle de polymère proche de la transition, cette fois-ci pour décrire la dynamique des chromosomes.

Dans l'ensemble, cette thèse fournit une base précieuse pour comprendre le comportement des polymères dans les systèmes biologiques et apporte des indications sur l'interprétation des données expérimentales relatives à la chromatine.

ABSTRACT

The primary objective of this thesis is to conduct a comprehensive examination of polymer models to establish reference results that enhance the analysis of experimental chromatin data. This research delves into both equilibrium and dynamic aspects, addressing fundamental questions about polymer behavior in biological systems.

In the first part of the thesis, the focus is on equilibrium properties, providing a theoretical framework for the study of sequential FISH data. This includes introducing essential concepts of equilibrium polymer physics, such as Gaussian chains. It also explores the concept of Rouse modes in a new perspective, connecting it to signal analysis techniques. Computer simulations of self-interacting polymers are performed and analyzed, leading to the development of a spectral-based order parameter that characterizes polymer folding states, up to the establishment of a precise phase diagram for the coil-globule phase transition.

The spectral approach is finally applied to sequential FISH data with two distinct results. Firstly, it detects a spectral signature indicating critical polymer behavior, thus validating the usefulness of simple polymer models in chromatin analysis. Secondly, thanks to the implementation of a specific data analysis pipeline, it reveals previously unidentified loops in the same data, offering new insights into the study of the fine chromatin organization.

The second part of the thesis shifts attention to dynamical aspects of the coil-globule phase transition. This part begins by introducing stochastic dynamics and polymer dynamics concepts, including the Langevin equation and the (dynamical) Rouse model. Extensive Langevin dynamics simulations of self-interacting polymers are presented, along with an *interacting Rouse model* that accounts for the coil-globule phase transition and excluded volume effects. The introduction of topological constraints in simulations highlights the robustness of Rouse dynamics, which are only significantly altered in dense globule phases.

In the final chapter, the acquired knowledge is applied to explain seemingly paradoxical data from a recent study of *Drosophila* chromatin dynamics, where again we find indications toward the pertinence of a polymer model at criticality, this time in describing chromosome dynamics.

Overall, this thesis provides a valuable foundation for understanding polymer behavior in biological systems and offers insights into interpreting experimental chromatin data.

List of publications

1. *Assessing the polymer coil-globule state from the very first spectral modes.*
   by Timothy Földes, Antony Lesage and Maria Barbi
   Physical Review Letters 127.27 (2021): 277801
   also available at hal.science.

2. *Spectral-based detection of chromatin loops in multiplexed super-resolution FISH data*
   by Michaël Liefsoens, Timothy Földes and Maria Barbi
   submitted to *Nature Communications* and available at arXiv.

3. We are currently in the process of writing an article incorporating the findings of chapters 6 and 7.

4. We plan to write a paper, or comment, on the findings of chapter 8.

Layout

Throughout the thesis, I'll be using colored boxes for specific content:

Chapter outline

Useful results from previous works

New results of the thesis

Demonstrations, additional calculations

Published/submitted/in preparation articles

REMERCIEMENTS

En premier lieu, j'aimerais remercier Daniel Jost et John Marko, qui ont accepté de rapporter mon manuscrit, et pour leurs critiques constructives, qui ont participé à améliorer la qualité finale du manuscrit. Je remercie également chaleureusement les membres examinateurs de mon jury, Jean-Louis Barrat, Antoine Coulon, Marie Jardat et Judith Miné-Hattab pour leur bienveillance et leurs questions pertinentes lors de la soutenance.

Ensuite, je souhaite exprimer ma gratitude envers l'équipe chromatine du LPTMC qui m'a accueilli au début de ma thèse: Maria, Jean-Marc, Annick, Vincent et Antony.

En particulier, je souhaite remercier Maria, d'abord, m'avoir fait confiance et embauché pour cette thèse. Mais surtout, d'avoir été une directrice de thèse parfaite, suscitant la jalousie de tous mes amis doctorants. Merci d'avoir été disponible presque 24/7 pour discuter de science, d'avoir été un soutien moral inconditionnel et pour ton aide opérationnelle dans l'écriture de mon manuscrit et pour toutes les tâches administrative. Je garderai toujours un excellent souvenir des heures passées devant le tableau blanc à réfléchir au mouvement des polymères, souvent juste pour le plaisir!

Je remercie également Jean-Marc pour m'avoir transmis son savoir en physique des polymères et en biologie du noyau, mais également pour nos longues discussions amicales qui vont certainement me manquer! Et bien sûr, un immense merci pour les viennoiseries matinales qui ont accompagnées mon café tous les jours pendant trois ans!

Ensuite, j'aimerais exprimer ma reconnaissance envers les jeunes du LPTMC pour l'ambiance conviviale: Jérèm, Brieuc, Lucjia, Marc, Saverio, Leo, Louise, Anna, Julien, Theotim, Aurélien, David, Davide et Nicolas. Une mention spéciale à Brieuc et Jeremie pour avoir respectivement tenté et réussi à me faire commencer le sport.

Je voudrais également remercier les permanents du groupe cantine pour avoir déjeuné avec les étudiants chaque midi, ce qui nous a permis de nous sentir pleinement intégrés au laboratoire : Jean-Noël, Rémi, Bertrand, Olivier, Nicolas (Dupuis et Sator), Maxim et Pascal. Merci pour les conseils avisés qui m'ont aidé à surmonter les aléas et difficultés de la thèse. Et une mention particulière à Pascal pour m'avoir aidé avec mes soucis informatiques.

Je souhaite également remercier ma famille. Mes deux frères Theo et Stan pour les séances d'escalade à Antrebloc qui ont contribué à maintenir l'équilibre travail loisir. Je suis également très reconnaissant envers mes parents, pour leur amour inconditionel, évidemmment, mais également de m'avoir assuré un cadre stable me permettant d'effectuer des études longues. Votre enthousiasme, à tous, pour mes accomplissements académiques a été un véritable moteur pour ma réussite.

Je voudrais aussi remercier certaines personnes qui ont été importantes dans ma formation de chercheur: Charles Simon, mon professeur de physique de première, et Marc Rousselle, qui m'ont tous les deux servi de modèle et de mentor et m'ont donné la confiance nécessaire pour entreprendre des études scientifiques. Merci également à Christophe Grippon pour m'avoir justement réorienté vers l'université, et à Jean-Baptiste Fournier de m'avoir accordé une place dans le master IPCS.

Enfin, je voudrais remercier Oriane ♡ d'avoir partagé ma vie pendant ces trois années. Notre relation a eu un impact profond sur moi, au point que mon entourage a parfois du mal à me reconnaître. Merci d'avoir rendu mon quotidien plus heureux et simple, en étant toujours affectueuse et réconfortante, même dans les moments les plus tendus. Maria a mentionné à ma soutenance que j'étais généralement très gai au travail, et je pense que c'est en partie grâce à notre relation.

Contents

0

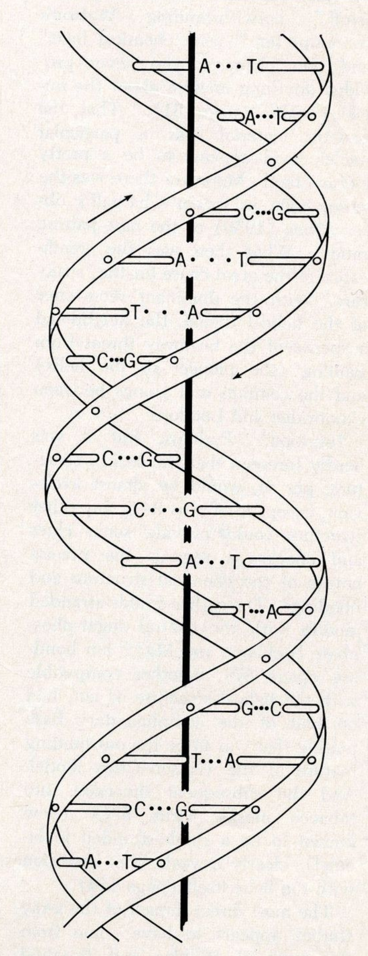
Introduction

In this introductory chapter, I aim to present key concepts of the field of chromatin organization, and layout the questions that will be tackled in the thesis.

In section 0.1, I present a brief overview of cell biology, with an emphasis on **chromatin organization** in the nucleus and its implications for cellular functions.

In section 0.2, I'll introduce the two main types of **experimental approaches** used to investigate chromatin organization. I will then discuss how the principles of polymer physics can be used to model chromatin and interpret this data. We'll see that polymer physics performs unevenly, depending on the complexity of the organism studied. Particular attention will be paid to the importance of the coil-globule phase transition in modeling chromatin structure and dynamics.

Finally, in section 0.3, I present the **objectives and outline** of this thesis.



DOUBLE HELIX. The two sugar-phosphate backbones twist about on the outside with the flat hydrogen-bonded base pairs forming the core. With these base pairs as the steps the structure resembles a helical staircase.

0.1 Chromatin

0.1.1 *The cell and the genome*

The elementary structural and functional unit of all living matter is the **cell**, a small compartment, delimited by a lipidic membrane, containing highly specialized molecules, whose properties were fine-tuned through the process of evolution, enabling the cell to self-sustain and replicate.

The instructions for the development, functioning, growth, and reproduction of the cell are encoded in a long linear **macromolecule**, or **polymer**, called deoxyribonucleic acid (DNA). DNA is made of two complementary strands, each of which is made of subunits, bound together by covalent bonds, called nucleotides. Each nucleotide can be one of four elementary molecules: cytosine (C), guanine (G), adenine (A), and thymine (T). Both strands are connected together by hydrogen bonds between facing nucleotides, forming the famous double helix. Two facing nucleotides always obey the pairing rule, T binds with A and G binds with C. The pairing rule implies that the sequence of one strand is completely determined by the other strand, resulting in both strands holding the exact same information. A couple of facing nucleotides are called a **base pair (bp)** and serve as the fundamental units of DNA.

Each living organism possesses a specific sequence of base pairs called its **genome**, stored in one or several DNA molecules called **chromosomes**.

In a process called transcription, specific parts of the genome, the **genes**, serve as templates to produce RNA strands, an almost identical molecule to DNA, where the thymine nucleotides are replaced with uracils. This single-stranded RNA is then **translated** into a **protein**, a linear macromolecule made of subunits called **amino acids**, that perform essentially all functions in the cell. The translation process involves the **genetic code**, a set of rules that specifies the correspondence between the nucleotide sequence of DNA, or equivalently RNA, and the amino acid sequence that constitutes proteins. Proteins differ from one another in their sequence of amino acids, which results in different proteins folding into a specific 3D structure. The 3D structure of a protein, in turn, determines its biological function.

There are two distinct cell types that differ by their internal structures: **Prokaryotes**, which are single-cell organisms, typically bacteria and archaea. They are characterized by the absence of internal compartments in their cells, with their genetic material typically found in an irregularly shaped region called the nucleoid.

The remaining living organisms, including mammals, insects, plants, fungi, as well as unicellular organisms such as yeast, are referred to as **eukaryotes**. Eukaryotes, are more complex, featuring a specialized compartment housing DNA, called the **nucleus**.

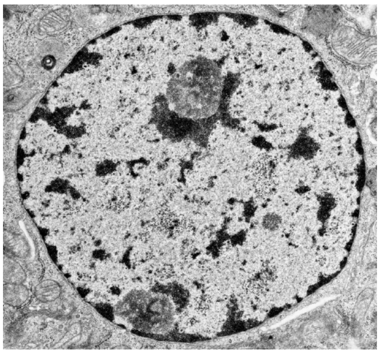


Figure 1: A human cell nucleus imaged by electron microscopy. Taken from Hystology@Yale.

The size of a genome differs dramatically among organisms. The organism with the smallest known genome is *Mycoplasma genitalium*, a bacteria, small in size (200-300 nm), responsible for a sexually transmissible disease. Its genome consists of approximately 580,070 base pairs stored in one circular chromosome. It codes for 482 proteins, making it the minimal known set of genes allowing for life to exist.

Conversely, the largest known genome belongs to the *Polychaos dubium*, a relatively large (0.5 to 1 mm) uni-cellular eukaryote, with a genome of 670 billion bp.

The human genome stands in between at 3 billion base pairs, distributed among 23 pairs of diploid chromosomes[1]. The genetic material is contained in a nucleus, which varies in size depending on the cell type, ranging from 5 to 10 μm (Figure 1).

When the entire human genome is fully stretched and chromosomes put head to toe, it spans approximately 2 meters in total[2]. Yet, a copy of the entire genome is stored in the micrometric nucleus of every single cell of the 37.2 trillion cells in the human body. This fact raises two central questions in molecular biology:

- First, given the length of the genome, how does it manage to fold into the small dimensions of the nucleus while retaining its biological

---

[1] Diploid cells have two homologous copies of each chromosome (i.e. carrying the same genes), one from the mother and one from the father.

[2] Each DNA base pair has a linear length of 0.34 nanometers, resulting in approximately, one meter per haploid set.

functions?[3]

- Second, considering every cell contains the same genetic information, how can we account for the diverseness of the observed cell types in a given organism?

Interestingly, the answers to these questions are linked, in that the specific spatial configuration adopted by DNA, following its folding inside the cell nucleus, is closely linked to the cell's morphology and biological function. This is due to the important role of DNA folding in the cell's **gene regulation strategies**.

To initiate transcription of a gene, several biological events must take place, including bringing together a number of elements in the vicinity of the gene start (in a region called *promoter*). These involve regulator proteins, called *transcription factors* (TF) which freely diffuse in the nucleus, a large protein complex called the mediator, and a distant DNA segment, called a **promoter**. Once these molecules have been brought together near the promoter, an RNA polymerase synthesizes a strand of RNA from the gene's sequence, which is then translated into a protein. From this elementary mechanism, different **gene expression regulation strategies** can be derived.

For example, the cell can deactivate a gene's expression by locally compacting DNA, making access to the genetic information impossible for TFs. Conversely, transcription can be stimulated by binding the enhancer to the promoter through connector proteins (CTCF), or by clustering several genes in physical proximity along with a high concentration of transcription factor, creating so-called transcription factories.

Compact regions of repressed chromatin, appearing as darker regions in Figure 1, are called **heterochromatin**. In contrast, decondensed regions where chromatin is typically actively transcribed, visible as bright regions in Figure 1, are called **euchromatin**.

These gene expression regulation strategies are, in part, implemented in the cell by tuning the physical distance between genomic elements in the nucleus, i.e. modulating the folding patterns of DNA in the nucleus. Hence, the specific spatial configuration of DNA, by regulating protein production levels, leads to different cell types emerging from the same genetic material.

Although the spatial structure adopted by DNA, and its relation to cellular functions, has been an active field of research for decades, our understanding of nuclear organization remains incomplete, with previous certainties frequently being challenged by advances in experimental methods. In the following section, I provide a concise overview of the current knowledge in this field.

0.1.2 *Chromatin Organization*

The folded DNA, associated with the structural proteins that orchestrate its spatial organization in the nucleus, is called **chromatin**. During interphase, the longest phase of the cell cycle, during which genes

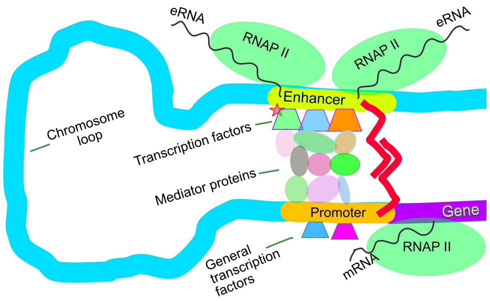


Figure 2: schematic representation of an enhancer-promoter interaction

---

[3] While 2 meters sounds like a huge amount of DNA, it is a slightly misleading number. Indeed, DNA has a very small cross-section (about 2 nm) and is consequently 1) highly flexible and 2) its total volume is relatively small compared to the volume of the nucleus, about 10% [3]. Hence, the challenge isn't so much to fit it into the nucleus, but rather to fold it in a fashion that retains biological processes such as gene expression, DNA replication, and repair.

are transcribed and DNA is replicated[4], chromatin is in a *decondensed* state and spreads out to fill the whole nucleus, allowing for translation (ribosomes) and replication (DNA polymerase) proteins to access the DNA.

→*Nucleosome*

Despite this homogeneous, space-filling appearance (see Figure 1), the internal organization of chromatin is very complex and involves multiple scales of partitioning.

At the smallest scale, DNA is wrapped around a **histone octamer**, a protein containing two copies each of 4 sub-proteins called H2A, H2B, H3, and H4. The DNA makes nearly two full turns around this octamer, wrapping about 147 base pairs. Histone H1 binds to the DNA as it enters and exits the nucleosome, helping to stabilize the structure. This structural unit is called the **nucleosome**. The length of the DNA segment connecting two nucleosomes, termed linker DNA, can range from 15 to 70 base pairs[5].

Histones H2A, H2B, H4, and H3 have intrinsically disordered parts, meaning parts that have no secondary structure, called histone tails. Histone tails are rich in the amino acid lysine (K in Figure 3). Lysines can be the target of biochemical modifications, called **epigenetic modifications**, which influence the physical properties of chromatin without altering the genetic code. Without going into the dauntingly complex details of the physics of epigenetics[6], we can say that modification of histone tails can essentially lead to two types of local physical alterations of chromatin properties:

(i) Change the *charge* of chromatin through the addition of an acetyl group (Acetylation), or a phosphate group (Phosphorylation) to a lysine of the histone tail. This can have the effect of either opening (+ charge) or compacting (- charge) chromatin at the location of the modification.

(ii) Recruit architectural *proteins*, such as proteins of the Polycomb group or of the family of heterochromatin protein 1 (HP1), which are known to compact chromatin locally.

These modifications of chromatin properties are one important pathway to the spatial organization of chromatin, yet still very largely unexplained in most organisms.

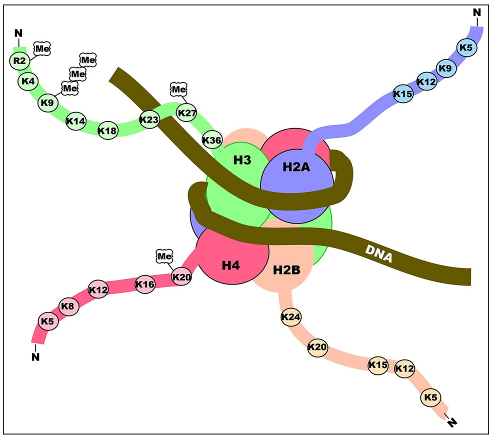


Figure 3: Schematic representation of a nucleosome. Taken from Wikimedia Commons. Courtesy to user MethylC5.

→*Large Scale Organization*

The small-scale nucleosome packing of DNA is highly universal, as it is shared by all living organisms, and fairly well documented. On the contrary, the higher orders of chromatin organization are species and cell-type dependent, and the general principles dictating their properties are still poorly understood. These higher-order structures can be defined as assemblages of nucleosomes that, on average, assume preferential physical proximity. As we've seen, epigenetic modifications can partly explain structural modifications of chromatin, yet it

[4] More precisely, the interphase is composed of three phases, during which the cell grows (G1), replicates its DNA (S) and prepares for mitosis (G2).

[5] F. Fatmaoui, P. Carrivain *et al.* Cryo-electron tomography and deep learning denoising reveal native chromatin landscapes of interphase nuclei. *bioRxiv*, 2022

[6] R. Cortini, M. Barbi *et al.* The physics of epigenetics. *Reviews of Modern Physics*, 88(2), Apr. 2016

is known that many other biological mechanisms are involved, notably *loop extrusion*, which will be discussed later in this thesis.

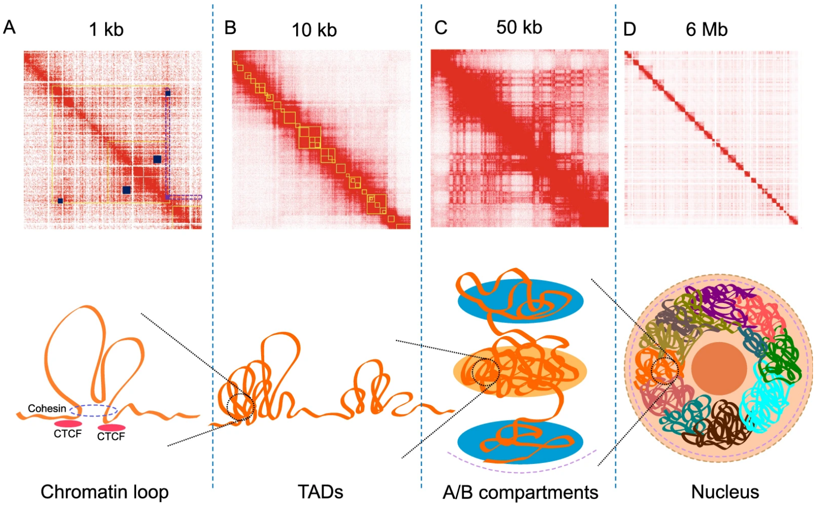


Figure 4: Illustration of genome architecture and the corresponding Hi-C interaction maps. Taken from [50].

The identification of higher order structures was made possible by advances in genomic analysis techniques, notably **Hi-C (high-throughput chromosome conformation capture)**, which captures the physical pair-wise contact frequency between all pairs of segments along the chromatin fiber. The resulting data, called a contact probability, or contact map, gives, for each pair of DNA sequences, an estimate of the probability to be found in close proximity within the cell.

With the advent of Hi-C, several structures have been confirmed or discovered. Figure 4 depicts Hi-C contact maps at different resolutions, illustrating the main known higher-order structures.

To start from the larger scale, in Figure 4-D a low-resolution contact map of the whole genome is shown. Square blocks appear along the main diagonal, indicating a higher probability of contact among segments within the block. Each block corresponds to the genomic coordinates of one chromosome. This pattern, called **chromosome territories**, indicates that regions within the same chromosome are more likely to interact with each other than with regions from different chromosomes.

By considering now a single chromosome, in Figure 4-C, a higher resolution Hi-C map reveals intra-chromosome interactions. The striking features at this scale are the off-diagonal blocks of enriched contacts, creating a sort of checkerboard pattern. The physical interpretation for the presence of these blocks is that chromosomes can be partitioned into two spatial compartments, coined **A and B compartments**, wherein greater interaction occurs within rather than between them, as originally described by Lieberman-Aiden et al. in 2009[7]. Regions within the A compartment are associated with actively expressing genes, and are typically less compact, euchromatic regions, while those in the B compartment typically contain expression-inactive genes, hence generally belong to heterochromatin regions.

[7] E. Lieberman-Aiden, N. L. van Berkum *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009

By enhancing again the resolution (see Figure 4-B), contacts within an A/B compartment can be distinguished. We notice that compartments are further organized into **topologically associating domains or TADs**[8], appearing as darker squares along the diagonal, which constitute isolated regions of chromatin that have augmented interactions within and remain largely independent of other chromatin regions. There is still debate on exactly what biological function these TADs have and how they are formed. One possible role of this organization is to make certain genes occupy separate regions in space, such that their individual activities are non-interfering[9].

Finally, the smallest known scale of chromosome organization involves chromatin **loops** (see Figure 4-A). The formation of such loops is generally attributed to the loop extrusion process[10], where proteins, condensins or cohesins, actively move along a DNA strand, grabbing and extruding a loop of DNA as they progress. The extrusion process stops when a protein, called CCCTC-binding protein (CTCF), is encountered, anchoring the base of the loop and shaping the looping structure. CTCF-mediated loops are thought to be a main player in the stabilization of TADs.

0.2 POLYMER PHYSICS AND CHROMATIN MODELING

The statistical properties of polymer **conformations** -the three-dimensional spatial arrangement of its molecules - due to their highly fluctuating nature, exhibit many universal characteristics. It is then reasonable to expect that the spatial organization and dynamics of chromatin should be greatly influenced by the polymeric nature of DNA. This prospect has captivated polymer physicists, who have devoted substantial efforts to explain different aspects of chromatin organization using staple **concepts and models of polymer physics**, the hope being that evolution would leverage the universal nature of polymer fluctuations to implement robust biomechanistic processes to organize chromatin.

The ultimate goal in chromatin modeling is, as usual in physics, to build a model, as simple as possible, that reproduces the available experimental data, and that allows for its interpretation, and the formulation of predictions. Due to the complexity of such models, analytical results are often out of reach, making computer simulations the preferred approach.

Regarding available data, two main categories of experimental approaches exist, and are employed to test polymer models. These approaches result in either **dynamic data** or **structural data**, respectively.

0.2.1 *Dynamic Data*

Using fluorescent probes that bind to specific chromatin loci, it is possible to track the position over time of, usually, one or two chromatin

[8] E. P. Nora, B. R. Lajoie *et al.* Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381–385, 2012

[9] Q. Szabo, F. Bantignies *et al.* Principles of genome folding into topologically associating domains. *Science Advances*, 5(4):eaaw1668, 2019

[10] E. Alipour and J. F. Marko. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic acids research*, 40(22):11202–11212, 2012

segments at once.

1. If a single segment is imaged, the diffusive properties of a **single monomer** of the model can be tested against those of the chromatin segment, generally by comparison of the **mean square displacements (MSD)** and the **two-time velocity correlation function**.

2. If two segments are imaged, in addition to the previous, other more informative observables can be computed, such as the **decorrelation time of the relative position of both segments**, or the MSD of one segment in the reference frame where the other segment is kept as the origin.

3. In addition, very recently, the **mechanical properties** of chromatin were tested by micromanipulation of ferromagnetic fluorescent bead allowing to probe the stress response of in-vivo chromatin[11].

Although some ongoing debate remains, notably regarding heterogeneities of the diffusion coefficient depending on the genomic location of the marker, the overwhelming majority of the community seems, at the moment, to agree that the simplest model for polymer dynamics, the **Rouse model**, well accounts for all the experimental data, regardless of the complexity of the organism. This reference model, which takes into account only the chain connectivity, will be extensively discussed in this thesis. The three canonical organisms in which Rouse dynamics were found, by ascending order of complexity, are yeast[12], *Drosophila*[13] and human[14]. The surprising fact that these three very different organisms display very similar dynamical behavior seems to point toward a universal polymer property playing an important role in chromatin organization.

0.2.2 *Population Averaged Structural data*

On the other hand, a more global observation of the spatial configuration of chromatin is possible, but at the price of averaging over a population of cells.

1. The main player in this regard is the aforementioned **Hi-C contact map**, which indirectly gauges the spatial conformation of the DNA strand in-vivo, at the population level. The Hi-C contact maps can directly be compared to contact frequency obtained from polymer simulations. Also, from the contact frequency, the averaged **contact probability** $P(s)$ of two segments as a function of their genomic distance s can be computed, both from experiments and simulations.

2. In 2016, an experimental study from Boettiger et al., using **super-resolved microscopy images** of *Drosophila* chromatin **domains**, allowed for a precise assessment of the scaling of a domain's physical 3D size with respect to its genomic length, that can also be compared to theoretical and numerical results.

[11] V. I. P. Keizer, S. Grosse-Holz *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605):489–495, 2022

[12] M. Socol, R. Wang *et al.* Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. *Nucleic Acids Research*, 47(12):6195–6207, May 2019

[13] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023

[14] V. I. P. Keizer, S. Grosse-Holz *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605):489–495, 2022

3. Finally, very recently, the most direct and information-dense observation of chromatin conformations was achieved by **sequential-FISH**. This method allows for the simultaneous localization of hundreds of chromatin segments in thousands of **fixed** cells at once, allowing, for the first time, a **direct visualization of the 3D path** of chromatin in-vivo. From this data set, essentially all the conformational information, up to the genomic resolution of the experiment (ranging from 1 kb to 1 Mb), is available, allowing for any ensemble averaged polymer physics observable to be computed, calling for innovative methods to be invented.

Unlike dynamic data, concerning structural observables, the success of simple polymer theory to account for experimental observations greatly depends on the complexity of the probed organism.

→ *Yeast*

Yeast being a unicellular organism, the uselessness of cell differentiation and hence of most gene regulation mechanisms, allows for a parsimonious organization of its chromatin, exhibiting essentially no higher-order structures. Consequently, several relatively simple **homogeneous polymer** simulations have been able to reproduce the organization of interphase yeast nuclei, in addition to explaining the main dynamical observables[15].

→ *Drosophila*

In the well-studied case of the *Drosophila*, which presents an intermediate level of complexity, recent investigations tend to indicate that polymer behavior is partially conserved, but with some significant changes that complexify the picture. A block copolymer model[16] and a string and binders model, that both, in a very similar fashion, implement **specific attractive interactions based on the epigenomic sequence**, were proposed. Fortunately, the epigenetics of *Drosophila* are relatively simple, involving essentially 3 different epigenetic states that influence significantly chromatin structure. These models successfully reproduce higher-order chromatin, displaying the importance of epigenetic-mediated interactions in the chromatin folding.

→ *Human*

In human chromatin, the influence of epigenetics on chromatin architecture is extremely complex, owing to an almost continuous spectrum of existing epigenetic states. An initial attempt, employing a simple homogeneous "**fractal globule**" model, was presented in the seminal Hi-C paper in 2009. This long-lived metastable state of homogeneous polymers was introduced 20 years before on theoretical grounds[17] for explaining the kinetics of collapse of homopolymers. Its claim to fame is to explain the scaling of the contact probability $P(s)$ while simultaneously providing a mechanism for the existence of chromosome territories[18]. Nevertheless, the fractal globule model is contested, as it has

[15] More precisely, except for telomeres and for the highly transcribed DNA in the nucleolus, yeast chromosomes behave as a polymer brush [83].

[16] D. Jost, P. Carrivain *et al.* Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*, 42(15):9553–9561, 08 2014

[17] A. Y. Grosberg, S. Nechaev *et al.* The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de Physique*, 49(12):2095–2100, 1988

[18] L. A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research*, 19:37–51, 2011

been shown, through computer simulations, to be unstable[19].

More elaborate models are now proposed, that allow for the reproduction of Hi-C maps but at the price of a huge amount of different **specific interactions**, whose parameters are fitted, often using machine learning techniques, directly on the Hi-C maps they intend to reproduce[20,21].

The fact that both human and *Drosophila* exhibit complex higher-order structures at the population level, while robustly exhibiting Rouse dynamics, is indeed a puzzling observation and has been described as a paradox by major actors in the community [22,23,24].

→ *Chromatin Lies at the Coil-globule Phase Transition*

A crucial insight, pertinent to the physical nature of chromatin, has been drawn from these simulation studies. Indeed, the value of the interaction parameters used to match the different structures of the Hi-C maps, appears to place the polymer at, or very close, to the **coil-globule phase transition**. The coil-globule phase transition of a self-attractive polymer describes the sudden collapse, from an extended conformation, called coil, to a compact conformation, called globule, due to changes in environmental conditions, such as a change in temperature or solvent quality.

This finding was confirmed by an in-depth analysis of Boettiger et al.'s super-resolution microscopy data of *Drosophila* chromatin domains. In this study, Lesage et al.[25] developed a theoretical framework allowing not only to fit the scaling of the average size of a domain, which was the approach adopted in the theoretical analysis of the original paper, but also to fit the **distributions** of domain sizes. This study found, with a high degree of certainty, that *Drosophila* chromatin domains, due to nucleosome-nucleosome interactions, indeed displayed the statistical behavior of self-attractive polymers at the coil-globule phase transition, confirming the hint initiated by the previous simulation studies.

The authors also elaborated on the biological relevance of such a property. By evolving into a critical state, chromatin can potentially change its conformation, hence its genetic activity, at a lower energetic cost, enhancing the **system's responsivity**. With this discovery, an intrinsic property of the polymer is found to be at the forefront of chromatin organization[26].

0.3 OBJECTIVES AND OUTLINE OF THE THESIS

The thrust of this thesis is to further challenge the **conjecture of the critical polymer model of chromatin** by testing it on the readily available experimental data, namely sequential FISH and dynamical loci tracking. Since comparing a model with experimental biological data is never trivial, our general approach was to produce, as a basis,

[19] R. D. Schram, G. T. Barkema *et al.* On the stability of fractal globules. *The Journal of Chemical Physics*, 138(22):224901, 06 2013

[20] A. Esposito, S. Bianco *et al.* Polymer physics reveals a combinatorial code linking 3d chromatin architecture to 1d chromatin states. *Cell Reports*, 38(13):110601, 2022

[21] M. D. Pierro, B. Zhang *et al.* Transferable model for chromosome architecture. *Proceedings of the National Academy of Sciences*, 113(43):12168–12173, 2016

[22] V. I. P. Keizer, S. Grosse-Holz *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605):489–495, 2022

[23] S. Grosse-Holz, A. Coulon *et al.* Scale-free models of chromosome structure, dynamics, and mechanics. *bioRxiv*, 2023

[24] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023

[25] A. Lesage, V. Dahirel *et al.* Polymer coil–globule phase transition is a universal folding principle of drosophila epigenetic domains. *Epigenetics & Chromatin*, 12(1), May 2019

[26] Not forgetting, of course, that the solvent, whose importance we are increasingly realizing today, is involved in maintaining the critical state. Solvent composition probably plays an important role in leveraging this critical state by operating these conformational changes.

a detailed study of both the **equilibrium** and the **dynamical** properties of **isolated** finite size self-interacting polymers, tailored for the analysis of both data types. This naturally divides the thesis into two main parts.

→ *First part*

The first part concerns solely **equilibrium properties**, in view of developing a theoretical framework to study the sequential FISH data. In doing so, we have developed an entirely new theoretical framework, based on **spectral analysis**, and especially adapted to these data sets.

- In the **1st chapter**, I briefly introduce basic concepts of equilibrium polymer physics. The intent of this chapter is twofold:
  (i) Introducing the well-known concept of ideal chain through the examples of the **freely jointed chain** and **Gaussian models**, while introducing typical equilibrium polymer physics observables.
  (ii) In a second part, I propose a new way of solving the Gaussian chain model, allowing for the introduction of a concept frequently used in the context of polymer *dynamics*, but not in that of equilibrium properties: the **Rouse modes**.

- In the **2nd chapter**, I make an original link between the concept of Rouse modes and well-known concepts of signal analysis. I show that the Rouse modes are a Fourier-like decomposition of the polymer chain considered as a stochastic signal, where the usual time domain of Fourier analysis is replaced by the curvilinear coordinates along the chain. Eventually, this allows for the interpretation of the fluctuations of the Rouse modes as a **power spectral density** of the polymer conformation.

- In the **3rd chapter**, I produce an extensive analysis of computer **simulated** self interacting polymers, using the power spectral density observable defined in chapter 2. I find that conformational changes occurring during the coil-globule phase transition only concern large-scale fluctuations, which the Rouse modes power spectral density are great at isolating. This allows for the definition of a spectral-based **order parameter** that **characterizes unequivocally the folding state** of a polymer. Based on this order parameter, I produce a thorough thermodynamic analysis of the model.

- In the **4th chapter**, I perform a spectral analysis of the **sequential FISH data** from Bintu et al., demonstrating that the data do indeed display the spectral signature of a critical interacting polymer. Finally, I close the chapter by extending the spectral approach to a different application: **detecting loops** in sequential fish data, and apply it to the same data. In this work, whose primary investigator is Michaël Liefsoens, a very talented intern student, we find previously undiscovered loops in the data.

→ *Second part*

The second part of the thesis is devoted to **dynamical aspects** of the **coil-globule phase transition**. To our great surprise, the dynamics of an isolated self-attracting polymer had never been investigated before. The aim of this section was therefore to carry out a systematic simulation study to catalog the dynamic properties of polymers at different values of the interaction parameter. With the aim of using dynamic data from one or two loci, we sought to find **dynamic markers** that could indicate the folding state of the polymer chain given the trajectory of only one or two monomers.

- In the **5th chapter**, I introduce elementary concepts of stochastic dynamics and polymer dynamics, namely the Langevin equation, and the **Rouse model**. In this context, the main dynamical observables will also be introduced.
- In the **6th chapter**, I present an extensive Langevin dynamics simulation study of self-interacting polymers. In an attempt to isolate the different contributions to the dynamics, I introduce the **interacting Rouse model**, a model that undergoes a coil-globule phase transition, accounts for excluded volume, but neglects topological constraints[27], an important physical mechanism that strongly influences the dynamics in very dense polymeric systems. Through the analysis of the Rouse modes of the chain, I show that Rouse dynamics is conserved at all stages of the collapse for this model.
- In the **7th chapter**, I introduce **topological constraints** in the simulations. By comparison, I establish the physical conditions, i.e. minimal chain length and interaction parameter, for topological constraints to have a visible effect on the polymer dynamics. I eventually show that Rouse dynamics are extremely robust, and are only altered in the dense globule phase.
- Finally, in **chapter 8**, I will attempt to make use of the acquired knowledge to explain **some seemingly paradoxical data** found in a recent study of *Drosophila* chromatin dynamics[28].

[27] I call here topological the constraints that prevent two polymer segments from going through each other.

[28] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023, [11].

PART I : SPECTRAL ANALYSIS OF POLYMER CONFORMATIONS

INTRODUCTION TO THE FIRST PART

The first part of my thesis concerns the equilibrium properties of real polymers. Monomers in a real polymer have finite volumes that cannot overlap and, through their interactions with the solvent and themselves, exhibit effective monomer-monomer attractive interactions. As this effective attraction increases, the polymer undergoes a phase transition, from a decondensed coil state to a dense globular state. The state of the polymer is generally identified by the scaling of its gyration radius with the number of monomer, meaning that one needs several data sets, at different chain lengths, to deduce a polymer's state.

The aim of this fist part of my work is to propose a phenomenological method, based on an original spectral analysis approach, to determine, instead, the folding state of a polymer from a set of conformations at a *single* polymer length. Within the new framework, I will propose an extensive simulation study of the coil-globule phase transition, introducing new observables and leading to some interesting results.

This spectral approach then allows me to develop a theoretical framework for studying sequential FISH data: in the last chapter of this first part, I will indeed be able to interpret the results of Bintu et al. [9] in the framework of a critical interacting polymer model, then to approach the question of loop detection from an original point of view, by introducing a new data analysis tool that effectively identifies loops in sequential FISH chromatin configurations datasets.

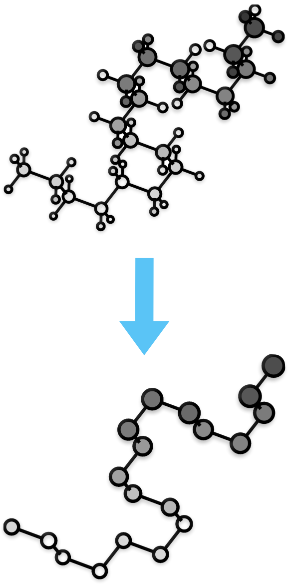
1

Equilibrium Polymer Physics

1.1 Ideal Chains, Real Chains

When in thermal equilibrium with a heat bath, a polymer visits a set of conformations determined by the interactions among all its constituent atoms and the surrounding solvent. Each conformation can be fully characterized by the position, velocity, and type of its atoms. The physics of such polymers could be investigated by evaluating the system's energy and computing macroscopic quantities using Boltzmann statistics. In practice, of course, keeping track of atomic details is impossible and one must resort to the study of coarse-grained models.

The relevance of **coarse-grained models** is based on the universality property of highly fluctuating systems. This states that, in the thermodynamic limit, the macroscopic properties of highly fluctuating systems do not depend on the microscopic details of the model. Formally, constructing a coarse-grained polymer amounts to, starting from the full description of a real polymer, regrouping microscopic degrees of freedom and redefining interactions between the remaining degrees of freedom in such a way that the statistics of the system remain unchanged. This procedure can be realized several times until the resulting model is simple enough to be simulated numerically or even solved analytically. Crucially, by coarse-graining two polymer models that differ at the microscopic scale, we sometimes find that they share the same large-scale behaviors. The set of all models sharing the same large-scale characteristics is called **class of universality**.



The "**ideal chain**" class groups together models that neglect all interactions between monomers that are not adjacent along the chain, thus isolating the most fundamental property of polymer molecules: chain connectivity. These models, however, allow conformations in which the polymer overlaps with itself, and are therefore sometimes referred to as *phantom* polymers. As we will see in the following section, in the absence of external forces, conformations of an ideal chain polymer can be represented by the outcome of a **simple random walk**. Consequently, the statistics of ideal chains is **Gaussian** at the macroscopic scale, regardless of the microscopic details of the model.

However, two segments of a real polymer cannot physically overlap, due to steric interactions between their monomers, called excluded volume in the context of polymer physics, and, consequently, generally do not behave like ideal chains. In chapter 3, we will study such **”real chains”** models. Nonetheless, ideal chain models prove very useful as they correctly describe polymers under conditions where excluded volume interactions are effectively screened. This may happen in a fluid, where a monomer interacts with both the solvent molecules and other monomers. The relative strength of these interactions determines whether monomers effectively attract or repel: in a ”good” solvent in which the polymer is very soluble (or equivalently at high temperature), monomers repel, while in a ”bad” solvent in which the polymer is insoluble (or at low temperature), they attract each other. However, at an intermediate, critical temperature called the **θ-temperature**, these effects compensate such that non-subsequent monomers effectively don't interact. In these conditions, polymers are predicted to behave like ideal chains. Similarly, in polymer solutions, when the polymer concentration exceeds a certain threshold, the excluded volume interaction is screened, and again the large-scale behavior of the chains becomes ideal.

In this chapter, I illustrate these concepts by introducing, in section 1.2, the **Freely Jointed Chain model** and in section 1.4 and section 1.3, the **Gaussian chain model** or **bead-spring model**. I'll define macroscopic observables typically used in polymer physics and compute them for both models, showcasing the simple scaling behavior of ideal chain models.

Subsequently, starting from subsection 1.4.2, I will present an unorthodox approach to solving the bead spring model that will lead me to introduce the **Rouse modes**, which will play a central role in the entire thesis. Finally, in section 1.5, I'll analyze and discuss their mathematical definition, allowing for a physical interpretation of each component as a specific geometrical feature of the conformation.

1.2 FREELY JOINTED CHAIN

Different approaches exist to model the statistical behavior of polymers. Discrete models represent conformations of complex linear macromolecules as a chain of $N+1$ 3D vectors. Assemblies of several atoms of the real polymer are represented by a vector $\vec{R}_n$, $n \in [0...N]$, pointing to the center of mass of the group of atoms. I note $\vec{r}_n = \vec{R}_{n+1} - \vec{R}_n$, with $n \in [0...N-1]$, the *bond* (or *link*) *vector* joining monomer n and $n+1$ (see e.g. Figure 1.1). In principle, bonds can have different lengths (we will have an example with the *Gaussian chain*) and relative orientations (these are for instance restricted in the *Freely Rotating Chain model*).

Here, I will start by focusing on the simpler model, the Freely

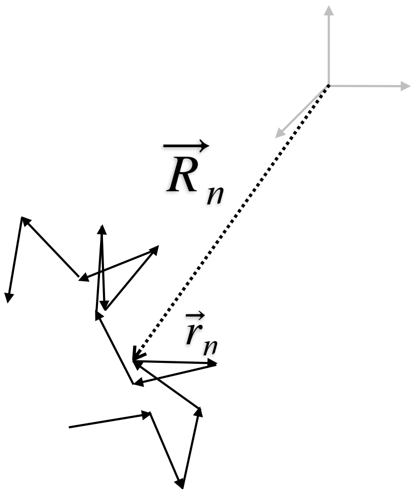


Figure 1.1: Schematic representation of a freely jointed chain.

Jointed Chain. In the **Freely Jointed Chain (FJC) model**, bond vectors have identical and fixed length $||\vec{r_n}|| = \ell$, and uncorrelated orientation:

$$\langle \vec{r}_n \cdot \vec{r}_m \rangle = \ell^2 \delta_{nm} \tag{1.1}$$

where δ_{nm} is the Kronecker delta. The sketch of Figure 1.1 refers to this model.

It is very interesting to note here the **analogy between the freely jointed chain and the random walk**, lying in their shared behavior of exploring a series of successive steps randomly. In a freely jointed chain, each segment or monomer is connected by flexible joints, allowing the chain to assume various conformations. Similarly, in a random walk, a particle takes random steps in different directions without any preferred path. From this mapping between the two models, it follows that all the well-known results for the random walk also apply to the FJC. I will, however, re-derive a few of them in the following.

Models that feature exponential bond orientation correlation functions are termed **semi-flexible** models. In contrast, those with delta-correlated[1] bonds are called **fully flexible** models. In this thesis will exclusively investigate fully flexible models

In any case, all these models belong to the same universality class, meaning that, with a little care in the choice of parameters, they all give the same large-scale behavior. The choice of FJC is therefore a reasonable one to start with.

1.2.1 *Measuring Polymer Size*

Because of the inherent flexibility of polymers, their typical conformation isn't a straight line. Instead, thermal fluctuations cause the chain to bend, resulting in the physical size of the polymer being generally less than its contour length. Studying the relation between the contour length (or number of monomers) and the physical size of a polymer has been the main focus of polymer physicists for the past century. Let's introduce the typical quantities useful for this matter.

→ *The end-to-end distance*

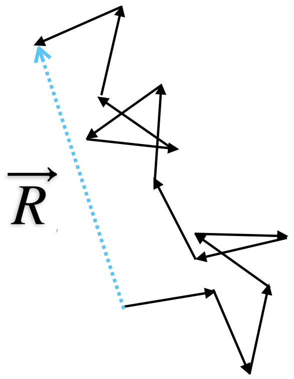
One possible way of measuring the size of a chain is its mean-square end-to-end distance[2]. We therefore introduce the **end-to-end vector** $\vec{R}$ (see again Figure 1.1)

$$\vec{R} = \vec{R_N} - \vec{R_0} = \sum_{n=0}^{N-1} \vec{r_n}. \tag{1.2}$$

We may immediately note that, by isotropy,

$$\langle \vec{R} \rangle = 0. \tag{1.3}$$

The **mean-square end-to-end distance** $\langle R^2 \rangle$ is then defined as the average of the squared amplitude of $\vec{R}$



[1] It is also worth noting that the assumption of random bond orientations implies as a hidden assumption that their length b is sufficiently large. Indeed, it is possible to assume that the correlation function of the tangent vector to a real macromolecule decreases (usually exponentially) with linear distance, thereby becoming random if we discretize sufficiently. It is also possible, of course, to take explicit account of the correlation of orientations, both in discrete models (as in the *Kratky-Porod* chain) or, for greater precision, in continuous models such as the *Worm Like Chain* (WLC), which models the polymer as a thin flexible rod.[76]

[2] It is common to define a (scalar and positive) *typical* end-to-end radius R as $\sqrt{\langle R^2 \rangle}$, this corresponding in other words to the standard deviation of the end-to-end vector.

$$\langle R^2 \rangle = \langle \vec{R} \cdot \vec{R} \rangle = \langle \sum_{n=0}^{N-1} \vec{r_n} \cdot \sum_{m=0}^{N-1} \vec{r_m} \rangle \tag{1.4}$$

$$= \sum_{i=0}^{N-1} \langle \vec{r_n}^2 \rangle + \sum_{n \neq m} \langle \vec{r_n} \cdot \vec{r_m} \rangle \tag{1.5}$$

In the case of the FJC, since $\langle \vec{r_n} \cdot \vec{r_m} \rangle = \ell^2 \delta_{nm}$, the rightmost term is zero and each element in the leftmost sum contributes ℓ^2, leading to the following scaling for $\langle R^2 \rangle$

$$\boxed{\langle R^2 \rangle = \ell^2 N} \tag{1.6}$$

We recognize here, as expected, the same results as for random walks: in both cases, the **root-mean-square displacement (RMSD)** of the system **grows as the square root of the number of steps** (either time steps or links), leading to a diffusive-like behavior with time, a characteristic commonly observed in systems undergoing Brownian motion. As we will see, this size scaling, proportional to N, is typical and defines the **ideal chain universality class**.

→*End-to-end Vector Probability Distribution Function*

As with random walks, more complete information on the distance between the last and first monomer can be obtained by calculating the probability distribution of the end-to-end vector. This is completely analogous to deriving, for the random walk, the probability distribution of moving a distance $\vec{R}$ after a number of steps N (the so-called propagator). However, I quickly retrieve this result here for completeness.

Let's compute the probability distribution that a FJC has a given end-to-end vector $\vec{R}$, noted $P(\vec{R})$. I start by writing $\vec{R}$ in terms of its spatial components R_x, R_y, R_z:

$$\vec{R} = R_x \vec{e_x} + R_y \vec{e_y} + R_z \vec{e_z} = \sum_{i=0}^{N-1} x_i \vec{e_x} + \sum_{i=0}^{N-1} y_i \vec{e_y} + \sum_{i=0}^{N-1} z_i \vec{e_z} \tag{1.7}$$

where we have introduced x_i, y_i and z_i the spatial components of $\vec{r_i}$ and have used Equation (1.2) for each component of $\vec{R}$. To simplify the computation, we can focus on one component, for instance, the x component.

Due to the homogeneous and fully flexible properties of the FJC model, the $\{\vec{r_i}\}$ are independent and identically distributed, implying that R_x is the sum of N independent and identically distributed random variables. In the large N limit, considering $\vec{r_i}$ has a finite mean and variance, the **central limit theorem** applies to $\vec{R}$ meaning its distribution converges to a Gaussian distribution, the parameters of which are provided by (1.3) and (1.6):

$$\langle R_x \rangle = 0 \tag{1.8}$$

$$\langle R_x^2 \rangle = \frac{\langle R^2 \rangle}{3} = \frac{N\ell^2}{3} \,. \tag{1.9}$$

This finally gives for the x component of $\vec{R}$:

$$P(R_x) = \left(\frac{3}{2\pi N b^2}\right)^{\frac{1}{2}} \exp\left(-\frac{3R_x^2}{2N\ell^2}\right) . \tag{1.10}$$

This immediately leads to

$$P(\vec{R}) = P(R_x)P(R_y)P(R_z) = \left(\frac{3}{2\pi N\ell^2}\right)^{\frac{3}{2}} e^{-\frac{3(R_x^2+R_y^2+R_z^2)}{2N\ell^2}}$$

and finally

$$\boxed{P(\vec{R}) = \left(\frac{3}{2\pi N\ell^2}\right)^{\frac{3}{2}} e^{-\frac{3R^2}{2N\ell^2}}} : \tag{1.11}$$

the full **probability distribution for the end-to-end vector $\vec{R}$** is given by a centered Gaussian with variance $\sigma^2 = N\ell^2/3$, in perfect analogy with the random walk model.

→*The internal end-to-end distance function*

Another observable that is typically studied in equilibrium polymer physics is **the *internal* end-to-end distance** $R^2(s)$. This quantity is defined as the average squared end-to-end distance of a subchain of linear length $s = |i-j|$, i.e.

$$\langle R^2(s)\rangle = \langle(\vec{R}_i - \vec{R}_j)^2\rangle = \sum_{n=i}^{j-1} \langle \vec{r}_n^{\,2}\rangle, \tag{1.12}$$

and is therefore a function of s. The observable $R^2(s)$ gives a richer information about the model than the end-to-end vector of the whole chain. It provides information about the **inner structure** of the chain.

The conformations of a FJC, like the Brownian particle path, have the property of being **statistically self-similar, or fractal**[3]. This means that the statistics of any subchain are identical, up to a scalar factor, to the statistics of the whole chain. Hence we can apply the scaling (1.6) for $\langle R^2\rangle$ to a subchain of length s yielding the following functional form for $R^2(s)$

$$\boxed{\langle R^2(s)\rangle = \ell^2 s}. \tag{1.13}$$

The **power-law** behavior of $R(s)$ evidences the self-similarity of the conformation: indeed, power laws are the only mathematical functions to possess the scale invariance property. [4].

We may note that, in the case of FJC, due to its self-similarity, $R(s)$ contains in fact no more information than R. However, more generally, $R(s)$ may be an indispensable tool for studying models that are not scale-invariant and bear an intrinsic length scale.

→*The gyration Radius*

A third quantity used to assess the physical size of the entire chain is the **mean-square radius of gyration** of the chain. It is defined as the spatial variance of the conformation

$$\langle R_g^2\rangle = \frac{1}{N}\sum_{n=0}^{N-1} \left\langle \left(\vec{R}_n - \vec{R}_{cm}\right)^2 \right\rangle \tag{1.14}$$

[3] The mathematical definition of a self-similar, or fractal phenomenon is that it behaves (statistically) the same when viewed at different scales on a dimension (space or time). This implies self-similarity at all scales, which does not, of course, apply strictly to any physical system. In the case of FJC, self-similarity must be indeed understood as being limited, on the small scale, by the length of the bonds, and, on the large scale, by the overall size of the polymer.

[4] Given a power law function

$$f(x) = ax^{\alpha},$$

scaling the argument x by a constant factor c only results in a proportional scaling of the function itself

$$f(cx) = ac^{\alpha}x^{\alpha} = c^{\alpha}f(x).$$

Where we have defined the **center of mass** of the polymer

$$\vec{R}_{cm} = \frac{1}{N}\sum_{n=0}^{N-1}\vec{R}_n. \tag{1.15}$$

From Equations (1.14) and (1.15), the following, alternative expression for the gyration radius, involving the internal mean square distance function $R(s)$, can be derived:

$$\langle R_g^2\rangle = \frac{1}{2N^2}\sum_{i,j=1}^{N} R(j-i). \tag{1.16}$$

From this expression, we see that $\langle R_g^2\rangle$ is the integral of the internal end-to-end distance function, and therefore gives a condensed, single-digit measure of the spatial distribution of monomers[5].From this equation, we can compute R_g^2 for the FJC using (1.13), as

$$\langle R_g^2\rangle = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|i-j| \tag{1.17}$$

$$= \frac{1}{2N^2}\,2\sum_{i=1}^{N} i(N-i) \tag{1.18}$$

$$= \frac{N(N+1)(N+2)}{6N^2} \sim \frac{\ell^2 N}{6} \tag{1.19}$$

We find that $\langle R_g^2\rangle$ and $\langle R^2\rangle$ share the same $\sim N$ scaling. This reflects of course the fact that both quantities are a measure of the characteristic size of the polymer.

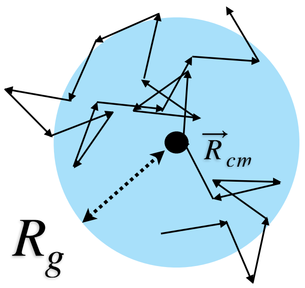


Figure 1.2: Center of mass $\vec{R}_{cm}$ and gyration radius R_g for a FJC conformation

1.3 From the Freely-Jointed Chain to the Gaussian Model

The ultimate goal of this chapter, in addition to presenting ideal chains and their properties, is to introduce the Rouse modes, which are defined as the normal modes of a polymer model, the Rouse model. The Rouse model, the simplest model for polymer *dynamics*, represents the chain as a sequence of beads, subjected to Brownian dynamics and connected by harmonic springs.

However, the Rouse model can also be introduced in the context of equilibrium polymer physics independently of the chain dynamics, and in this context, it is - surprisingly - generally called instead the Gaussian chain model.

Interestingly, the Gaussian chain model can be derived from a FJC in which large subsequent subchains are replaced by harmonic springs. Moreover, as we will see, this derivation takes us from an entropy-based model, with no energy difference between the conformations, to a mechanical model, where we can define a Hamiltonian.

As I shall show, there is an alternative derivation of the Rouse modes, fully independent of the dynamic aspects: the Rouse modes can be introduced by diagonalizing the system's Hamiltonian.

To begin with, I present the classical derivation of the Gaussian model from the FJC.

[5] Note, indeed, its close relation to the moment of inertia of the monomers ensemble.

1.3.1 *The Spring-Like Behavior of the FJC*

Polymeric materials, such as rubber, are known for their high elasticity. This behavior is due to the polymer's propensity to organize itself into random coils. If stretched into a straighter conformation, thermal fluctuations drive a polymer to return to its preferred coiled form, inducing a restoring force. For this reason, a long enough FJC can be seen as a harmonic spring, and both ends of the chain as interacting through a harmonic potential. To show this formally, we need to compute the free energy of the chain at fixed $\vec{R}$, which in turn depends on the end-to-end vector probability distribution function.

→*Free Energy and Harmonic Entropic Interaction Between Both Ends of the FJC*

From $P(\vec{R})$ we can calculate the **free energy** of the system. Since there is no interaction between the monomers, the internal energy of the model can be set to zero, and the only contribution to the free energy is from the **entropy** of the system.

$$S = k_B \ln(P(\vec{R})) = S_0 - \frac{3}{2} k_B T \frac{R^2}{N\ell^2}. \tag{1.20}$$

Where $S_0 = -3/2 \ln(\frac{3}{2\pi N\ell^2})$ comes from the normalization constant of $P(\vec{R})$. The free energy at constant $\vec{R}$ then reads

$$F(\vec{R}) = -TS = F_0 - k_B T \ln(P(\vec{R})) = F_0 + \frac{3}{2} k_B T \frac{R^2}{N\ell^2}. \tag{1.21}$$

Where $F_0 = TS_0$ is a constant with respect to $\vec{R}$.

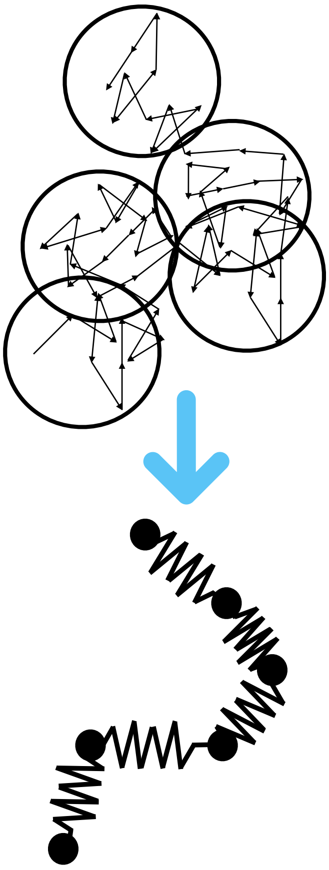
We see that the free energy is minimum for a zero end-to-end vector, and increases **quadratically** as the magnitude of the end-to-end vector grows in magnitude. A quadratic dependence of the interaction energy between two particles is called a **harmonic potential**. This means that, if we consider the system to be composed of only the two ends, forgetting about all other monomers between them, it behaves exactly like a harmonic spring. In other words, we can consider the free energy of the FJC at constant $\vec{R}$ as an **effective harmonic potential between both ends of the chain**

$$U_{\text{eff}}(\vec{R}_N - \vec{R}_0) = U_0 + \frac{3}{2} k_B T \frac{R^2}{N\ell^2} = U_0 + \frac{k}{2} \frac{R^2}{N\ell^2} \tag{1.22}$$

with (entropic) **spring constant**

$$k = \frac{3k_B T}{N\ell^2} \tag{1.23}$$

Furthermore, due to the fractal nature of the FJC, the same reasoning can be applied to any sufficiently long subchain of a FJC.



1.3.2 *Coarse-graining the FJC Into the Gaussian Model*

Now let's use the expression for the U_{eff} to derive another ideal chain model by **coarse-graining** the FJC.

Consider a FJC polymer of $N_{\text{FJC}}+1 \gg 1$ monomers with positions[6]

[6] I use cursive letters here to designate the FJC model, to distinguish it from the Gaussian chain, which will then be described again using the standard notation.

$\vec{\mathscr{R}}_0...\vec{\mathscr{R}}_{N_F}$ and bonding vectors $\vec{\mathscr{r}}_1...\vec{\mathscr{r}}_{N_F}$ of length ℓ.

The idea is to group the original degrees of freedom locally and redefine interactions while conserving the statistics of the chain. We start by grouping degrees of freedom. To do so, we divide the original polymer into N subchains of m monomers such that $m \gg 1$ and $N_{\mathrm{FJC}} = mN$.

We note the $N+1$ extremities of the subchains $\vec{R}_i = \vec{\mathscr{R}}_{m\times i}$ and the N bond vectors $\vec{r}_i = \vec{R}_i - \vec{R}_{i-1} = \sum_{k=m(i-1)}^{mi} \vec{\mathscr{r}}_{F_k}$. A conformation of the new model is equivalently represented by the set of vectors $\{\vec{R}_0...\vec{R}_N\}$ or by the bonding vectors $\{\vec{r}_0...\vec{r}_{N-1}\}$.

Now, let's define the nature of the new links between the new monomers. In the previous section, we have established that an entropic harmonic potential, of constant $k = \frac{3k_BT}{N\ell^2}$, acted between both ends of a FJC polymer of size $N \gg 1$. Applied to each polymer subchain, this suggests a harmonic potential between subsequent monomers of the coarse-grained model:

$$U_{i,i+1} = \frac{1}{2}\frac{3k_BT}{m\ell^2}(\vec{R}_{i+1} - \vec{R}_i)^2 = \frac{1}{2}k(\vec{R}_{i+1} - \vec{R}_i)^2 \tag{1.24}$$

where we have defined $k = 3k_BT/m\ell^2$ as an effective elastic constant.

1.4 THE GAUSSIAN CHAIN AND AN INTRODUCTION TO THE ROUSE MODES

1.4.1 *The Gaussian Chain*

The resulting model is a chain of $N+1$ monomers interacting through harmonic bonding potentials and is called the **Gaussian model** or **bead spring model**: subchains are modeled as identical springs of stiffness k, representing the entropic springs of the FJC. Modeling an entropic elasticity as a mechanical elasticity allows us to define a **Hamiltonian** for the model, which happens to be diagonalizable, making the model solvable analytically.

The energy of a configuration $\{\vec{R}_1...\vec{R}_N\}$, i.e the Hamiltonian of the model, is indeed:

$$\boxed{\mathscr{H}(\{\vec{R}_0...\vec{R}_N\}) = \sum_{i=0}^{N-1} \frac{1}{2}k(\vec{R}_{i+1} - \vec{R}_i)^2.} \tag{1.25}$$

Since the Gaussian model was derived from coarse-graining a FJC, we expect their size scaling to be identical. For good measure, let's compute $\langle \vec{R}^2 \rangle$. We remind its definition (1.5)

$$\langle R^2 \rangle = \sum_{i=0}^{N-1} \langle \vec{r_n}^2 \rangle + \sum_{n \neq m} \langle \vec{r_n} \cdot \vec{r_m} \rangle \tag{1.26}$$

Since no diagonal terms appear in $\mathscr{H}$, the bond cross-correlation is zero and the rightmost term $\sum_{n\neq m} \langle \vec{r_n} \cdot \vec{r_m} \rangle$ vanishes.

To finish the calculation, we must compute the **average bond size**, $b^2 = \langle \vec{r_n}^2 \rangle$. Interestingly, for a particular value of k, the average size of the bond is completely determined by the thermal fluctuations of the system. Indeed, the rest length of each harmonic bond is actually 0, but beads are kept apart on average at a distance b by the thermal energy stored in each bond at equilibrium.

The exact value of b is fixed by the **equipartition theorem**, stating that each bond $\vec{r}_i = \vec{R}_{i+1} - \vec{R}_i$, given it appears squared in the Hamiltonian (1.25), is distributed an equal amount of energy at equilibrium, yielding:

$$\boxed{\langle \vec{r_n}^2 \rangle = b^2 = \frac{3k_BT}{k}.} \tag{1.27}$$

Eventually leading to the following scaling for the mean square end-to-end distance

$$\boxed{\langle R^2 \rangle = b^2 N} \tag{1.28}$$

Which coincides with the (1.6).

As for R_g and $R(s)$, we also find the same formulas as for FJC, for which I am not going to detail the calculation, since they follow the same logic as for the FJC.

$$\langle R_g^2 \rangle \sim \frac{b^2 N}{6} \tag{1.29}$$

$$\langle R^2(s) \rangle \sim b^2 s \tag{1.30}$$

In any case, we retrieve the typical $\sim N$ scaling, confirming the universal behavior of ideal chains.

1.4.2 *The Rouse Modes as Eigenmodes of the Gaussian Chain Model*

The theory I've presented so far is well established and can be found in several textbooks, some of which I've used as inspiration[7,8,9]. Conversely, the derivation and interpretation of Rouse modes in the context of the Gaussian model developed at the end of this chapter is, to my knowledge, a original approach.

1.4.3 *Matrix Representation Of The Gaussian Chain Hamiltonian And Associated Eigenvalue Problem*

Due to the harmonic nature of the Gaussian chain bonds, its Hamiltonian is a quadratic form in the monomer positions, and can be expressed in matrix form as

$$\mathcal{H}(\{\vec{R}_0, \ldots, \vec{R}_N\}) = \frac{k}{2} \begin{pmatrix} \vec{R}_0 \\ \vdots \\ \vec{R}_N \end{pmatrix}^T \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \vec{R}_0 \\ \vdots \\ \vec{R}_N \end{pmatrix} \tag{1.31}$$

$$= \frac{k}{2} \mathbf{R}^T A \mathbf{R}$$

---

[7] A. Khokhlov. *Statistical Physics of Macromolecules*. AIP series in polymers and complex materials. AIP Press, 1994

[8] I. Teraoka. *Polymer Solutions, An Introduction To Physical Properties*. Wiley-Interscience, 2002

[9] J. Van Der Maarel. *Introduction To Biopolymer Physics*. World Scientific Publishing Company, 2007

[10] I use capital letters for square matrices and bold capital letters for arrays of 3D vectors, which are formally $N \times 3$ matrices

Since the matrix A^{10} is both real and symmetric, it is diagonalizable. Consequently, we can find $N+1$ eigenvalues $\{\lambda_p \,|\, p \in [0...N]\}$ and the same amount of **orthogonal** eigenvectors $\boldsymbol{\varphi}_p$ such that the following holds

$$A\boldsymbol{\varphi}_p = \lambda_p \boldsymbol{\varphi}_p \tag{1.32}$$

This amounts to find a diagonal matrix D, whose diagonal elements are the eigenvalues, and a **transfer matrix** Φ, whose columns are the eigenvectors

$$A = \Phi^T D \Phi \tag{1.33}$$

where

$$\Phi = \begin{pmatrix} - & \boldsymbol{\varphi}_1^T & - \\ - & \boldsymbol{\varphi}_2^T & - \\ \vdots & \vdots & \vdots \\ - & \boldsymbol{\varphi}_N^T & - \end{pmatrix} \text{ and } D = \begin{pmatrix} \lambda_0 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N \end{pmatrix} \tag{1.34}$$

Diagonalizing $\mathscr{H}$ amounts to operating a change of position variables in phase space. This can be seen by substituting (1.33) into (1.31), giving

$$\mathscr{H}(\{\vec{R}_0, \dots, \vec{R}_N\}) = \frac{k}{2}\mathbf{R}^T(\Phi^T D \Phi)\mathbf{R} \tag{1.35}$$

$$= \frac{k}{2}(\Phi\mathbf{R})^T D(\Phi\mathbf{R}) \tag{1.36}$$

From this expression, a new set of variables $\mathbf{X} = (\vec{X}_0 \dots \vec{X}_N)$ can be defined as follows

$$\mathbf{X} = \Phi\mathbf{R} \tag{1.37}$$

and the Hamiltonian, in matrix notation, becomes

$$\mathscr{H}(\{\vec{X}_0, \dots, \vec{X}_N\}) = \mathbf{X}^T D \mathbf{X}. \tag{1.38}$$

By abandoning the matrix notation, the pth new variables $\vec{X}_p$ can be expressed as the projection of the position vectors $\vec{R}_n$ onto the basis vectors $\boldsymbol{\varphi}_p$

$$\boxed{\vec{X}_p = \boldsymbol{\varphi}_p \cdot \mathbf{R} = \sum_{n=0}^{N} (\boldsymbol{\varphi}_p)_n \vec{R}_n} \tag{1.39}$$

and the Hamiltonian as the sum of the squares of the transformed variable, each multiplied by its corresponding eigenvalues

$$\boxed{\mathscr{H}(\{\vec{X}_0, \dots, \vec{X}_N\}) = \sum_{p=0}^{N} \frac{k}{2}\lambda_p \vec{X}_p^2}. \tag{1.40}$$

1.4.4 *Eigenvectors And Eigenvalues Of The Gaussian Chain Hamiltonian*

In the following, we give the explicit formula for both the eigenvalues and eigenvectors without going through the calculation.

The eigenvectors $\{\boldsymbol{\varphi}_p \,|\, p \in [0, N]\}$ are given by:

$$\boxed{\boldsymbol{\varphi}_p = \begin{pmatrix} \frac{1}{(N+1)}\cos(\frac{p\pi}{N+1}(0+1/2)) \\ \frac{1}{(N+1)}\cos(\frac{p\pi}{N+1}(1+1/2)) \\ \vdots \\ \frac{1}{(N+1)}\cos(\frac{p\pi}{N+1}(N+1/2)) \end{pmatrix} \text{ for } 0 \le p \le N,} \tag{1.41}$$

And the elements of the $N \times N$ **transfer matrix** Φ can be written, for $p > 0$

$$\Phi_{np} = \frac{1}{(N+1)}\cos(\frac{p\pi}{N+1}(n+1/2)) \tag{1.42}$$

These eigenvectors are, by construction, orthogonal but not normalized since $\|\boldsymbol{\varphi}_p\| = \frac{1}{N+1}$. This entails that the matrix Φ doesn't satisfy the usual orthogonality identity $\Phi^T\Phi = I$, but instead satisfies :

$$\Phi^T\Phi \equiv \Delta = \begin{pmatrix} \frac{1}{N+1} & 0 & \cdots & 0 \\ 0 & \frac{2}{N+1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{2}{N+1} \end{pmatrix} \tag{1.43}$$

which is equivalent to :

$$\Phi^{-1} = \Lambda^{-1}\Phi^T \tag{1.44}$$

And implies that the usual diagonal form of A, (1.33) in fact writes

$$A = \Phi D \Phi^{-1} = \Delta^{-1}\Phi D \Phi^T \tag{1.45}$$

This choice of normalization will be justified later (section 1.5).

The corresponding eigenvalues are

$$\boxed{\lambda_p = 4\sin^2\left(\frac{p\pi}{2(N+1)}\right)}. \tag{1.46}$$

Their dependence on p is shown in Figure 1.3.

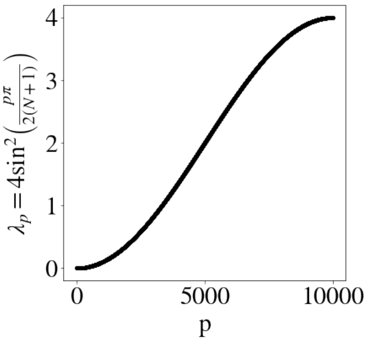


Figure 1.3: The eigenvalues λ_p of the Hamiltonian (1.31), according equation (1.46).

explicit Rouse modes expression

We can now write an explicit expression for the transformed variables using (1.39) and the formula for the eigenmodes

$$\boxed{\vec{X}_p = \sum_{n=0}^{N}(\boldsymbol{\varphi}_p)_n R_n = \frac{1}{N+1}\sum_{n=0}^{N}\vec{R}_n\cos\left(\frac{p\pi}{N+1}(n+1/2)\right)} \tag{1.47}$$

and from (1.44), we can retrieve the spatial coordinates from the Rouse modes:

$$\boxed{\vec{R}_n = \vec{X}_0 + 2\sum_{p=1}^{N}\vec{X}_p\cos\left(\frac{p\pi}{N+1}(n+\tfrac{1}{2})\right)}. \tag{1.48}$$

Transformed Rouse Hamiltonian

Considering our normalization choice and noticing that $\lambda_0 = 0$, the Gaussian Hamiltonian can be written in its diagonal form:

$$\boxed{\mathcal{H}(\{\vec{X}_0, \ldots, \vec{X}_N\}) = \sum_{i=1}^{N} \varepsilon_p \; \vec{X}_p^{\;2}} \qquad (1.49)$$

where we have defined

$$\varepsilon_p = k(N+1)\;\lambda_p = k(N+1)\;4\,\sin^2\left(\frac{p\pi}{2(N+1)}\right) \qquad (1.50)$$

1.5 THE ROUSE MODES

The new variables $\vec{X}_p$ are called **the Rouse modes** of the polymer. They play a crucial role in the study of polymer dynamics, serving as a mathematical tool to solve the dynamical version of the Gaussian model, known as the Rouse model. Indeed, one can use the autocorrelation function of $\vec{X}_p$ to compute several dynamical observables, such as the single monomer mean square displacement or the end-to-end vector correlation function. In the second part of this thesis, I will employ Rouse modes in their conventional use to analyze the dynamics of real chains. However, in this chapter, I'll propose an original interpretation of their mean squared average amplitude, allowing for a novel characterization of the equilibrium conformations of the Gaussian model and most importantly, in chapter 3, of interacting polymers.

1.5.1 *The Rouse Modes Describe Geometric Features of the Chain*

To further apprehend the information contained in the Rouse modes, let's analyze their mathematical definition. For a conformation $(\vec{R}_0, \vec{R}_1 \ldots \vec{R}_N)$, mode p writes:

$$\vec{X}_p = \sum_{n=0}^{N} u_p(n) R_n = \frac{1}{N+1} \sum_{n=0}^{N} \vec{R}_n \cos\left(\frac{p\pi}{N+1}(n+1/2)\right) \qquad (1.51)$$

where we have defined $u_p(n) = \cos\left(\frac{p\pi}{N+1}(n+1/2)\right)$

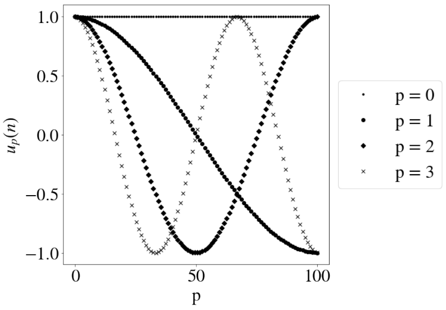


Figure 1.4: The first 4 basis functions $u_p(n)$.

Let's start with the zeroth Rouse mode, which, thanks to the normalization choice, corresponds exactly to the **center of mass** of the conformation

$$\vec{X}_0 = \frac{1}{N+1} \sum_{n=0}^{N} x_n = \overline{X}. \qquad (1.52)$$

The remaining modes contain the information of spatial features of the chain, on length scales decreasing as the mode number p, or, equivalently, the frequency of u_p increases. To understand why, following

Ref.[11], let's consider a modified version of $u_p(n)$ that is +1 when it is positive and -1 when the basis function is negative, i.e.

$$\widetilde{u}_p(n) = \mathrm{sgn}(u_p(n)) \tag{1.53}$$

and the modified Rouse modes:

$$\widetilde{\vec{X}}_p = \frac{1}{N+1} \sum_{n=0}^{N} R_{x_n} \widetilde{u}_p(n). \tag{1.54}$$

which, even if the actual coefficients in $\vec{X}_p$ change smoothly, should still behave roughly as mode p. Let's write the expression of the first modified mode:

$$\widetilde{\vec{X}}_1 = \frac{1}{N} \sum_{n=0}^{\frac{N}{2}} \vec{R}_n - \frac{1}{N} \sum_{n=\frac{N}{2}}^{N} \vec{R}_n = \frac{1}{2}(R_{\mathrm{cm},\ 1/2} - R_{\mathrm{cm},\ 2/2}) \tag{1.55}$$

where, if we partition our polymer into p segments, $R_{\mathrm{cm},\ \mathrm{q/p}}$ is the vector pointing to the center of mass (CM) of the qth segment. Thus $\widetilde{\vec{X}}_1$ is simply the vector extending from the center of mass of the first half of the chain to the center of mass of the second half of the chain. In fact, the smoothness of $\vec{X}_1$ makes it closer to the end-to-end vector. Consequently, $\vec{X}_1$ encapsulates the **orientation of the chain**, which is a feature that involves the whole chain and we thus say that it describes the polymer on the scale N, while any finer details are completely overlooked.

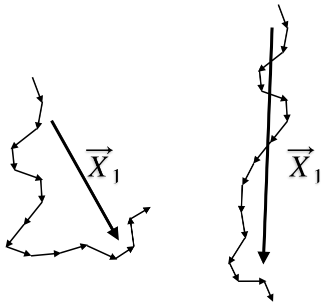


Figure 1.5: Illustrative examples of the first mode vector, $\vec{X}_1$.

Using the same reasoning for the second mode we find

$$\begin{aligned}\widetilde{\vec{X}}_2 &= \frac{1}{4}\left(\frac{4}{N}\sum_{n=0}^{\frac{N}{4}} \vec{R}_n - \frac{4}{N}\sum_{n=\frac{N}{4}}^{\frac{2N}{4}} \vec{R}_n - \frac{4}{N}\sum_{n=\frac{2N}{4}}^{\frac{3N}{4}} \vec{R}_n + \frac{4}{N}\sum_{n=\frac{3N}{4}}^{\frac{4N}{4}} \vec{R}_n\right) \\ &= \frac{1}{4}\left((R_{\mathrm{cm},\ 1/4} - R_{\mathrm{cm},\ 2/4}) + (R_{\mathrm{cm},\ 4/4} - R_{\mathrm{cm},\ 3/4})\right)\end{aligned}$$

Thus $\widetilde{\vec{X}}_2$ is the sum of the two vectors (green vectors in Figure 1.6), one from the CM of the second to the CM second quarters and the other from the CM of the third to the fourth.

From Figure 1.6, where two conformations with different values of $\vec{X}_2$ are displayed, it's easy to see that the second Rouse mode gives an idea of the **overall curvature of the chain** while remaining insensitive to curvature on scales smaller than $N/2$. Thus $\vec{X}_2$ isolates chain features with spatial scale $N/2$.

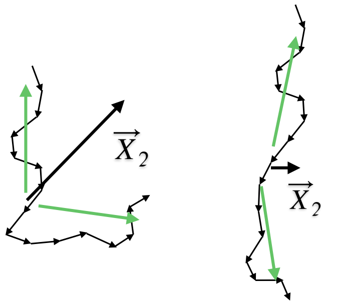


Figure 1.6: Illustrative examples of $\vec{X}_2$, which is the sum of the two green vectors.

Geometrical interpretation of the Rouse modes

Applying the same line of thought, we observe that higher-order modes exhibit a similar pattern and reveal progressively finer details of the chain. The pth mode is the sum of p vectors, each pointing from the center of mass of one subchain of size N/p to the next. Hence the pth mode describes and isolates a geometric

---

[11] I. Teraoka. *Polymer Solutions, An Introduction To Physical Properties.* Wiley-Interscience, 2002

feature of the chain on a scale N/p.

1.5.2 *Equipartition of Energy in the Rouse Modes*

Given the expression for the total energy (1.49), we can interpret the quantity

$$E_p = \varepsilon_p \vec{X_p}^2 \tag{1.56}$$

as the contribution of the pth mode to the total energy of a conformation. Conversely, by inverting the equation, we see that by exciting mode p with an energy E the squared amplitude of mode p will be:

$$\vec{X_p}^2 = \frac{E}{\varepsilon_p} \tag{1.57}$$

In essence, when two modes receive the same energy E, the resulting amplitude observed for each mode will be different and is determined by the coefficient ε_p^{-1}. In other words, the coefficient ε_p gives the energy required to modify the amplitude of mode p. Given the p-dependence of ε_p (equation (1.50)), we see that exciting low p modes requires less energy than exciting high p modes.

From equation (1.40), we notice that the quantity $\epsilon_p X_p^2$ enters squared in the Gaussian Hamiltonian. By virtue of the **equipartition theorem**, the average mode amplitudes must satisfy

$$\langle X_p^2 \rangle = \frac{3}{2} k_B T \frac{1}{\varepsilon_p}. \tag{1.58}$$

Rouse modes equilibrium fluctuations
for the Gaussian model

Thanks to (1.50) we can express the average square amplitude or **equilibrium fluctuations of the Rouse modes** as

$$\boxed{\langle X_p^2 \rangle = \frac{3}{8} \frac{k_B T}{k(N+1)} \frac{1}{\sin^2\left(\frac{p\pi}{2(N+1)}\right)}} \tag{1.59}$$

Note that, since $\langle \vec{X_p} \rangle = 0$, $\langle X_p^2 \rangle$ can be seen as the modes fluctuations. Finally, we can expand the sinus to the first order in $p \ll N$ and use the formula for the average inter-monomer distance $b^2 = \frac{3k_BT}{k}$ to get the **scaling behavior** for small enough p:

$$\boxed{\langle X_p^2 \rangle \approx \frac{b^2}{2\pi^2} \frac{N+1}{p^2}} \sim \frac{1}{p^2}. \tag{1.60}$$

As we will see, this p^{-2} scaling law will play a very important role in the following, as a reference corresponding to the ideal chain model.

1.6 Conclusion / take-home messages

In this chapter, I introduced the FJC and showed that, on large scales (and for low forces), to minimize entropy, the system behaves as a harmonic spring. Using this fact and the coarse-graining procedure, I derived from the FJC an even simpler model, the Gaussian chain, which assumes harmonic springs as bonding potentials between monomers. By calculating $\langle R^2 \rangle$ (or, equivalently, R_g^2) I showed that both models exhibit the same size scaling behavior

$$\langle R^2 \rangle \sim N$$

which characterizes the ideal chain universality class.

I then introduced the normal modes of the Gaussian chain model, the Rouse modes:

$$\vec{X}_p = \frac{1}{N+1} \sum_{n=0}^{N} \vec{R}_n \cos\left(\frac{p\pi}{N+1}(n+1/2)\right).$$

I gave a physical interpretation of this mathematical expression, showing that each mode $\vec{X}_p$ can be seen as a geometrical feature of the chain, on a linear length N/p.

The study of the partition of the ideal polymer's energy into different Rouse modes, reveals a power law decay of their equilibrium fluctuations as a function of the mode's number:

$$\langle X_p^2 \rangle \sim \frac{N}{p^2}.$$

It turns out that this functional form is directly related to the chain structure and can therefore provide a very useful tool for polymer conformation analysis. To make further progress, we need a theoretical framework to make sense of the expression for fluctuations of the Rouse modes (1.59). Conveniently, there exists a field dedicated to this subject: the theory of signal processing, which will be the focus of the next section.

2

Rouse Modes Fluctuations as a Power Spectral Density

Previously, we've established that ideal polymers, by the fluctuating nature of their conformations, can mathematically be described as the outcome of a simple random walk. In fact, a more general statement is true for any polymer model, namely that a polymer conformation can be described as a sequence of ordered random variables, i.e. a stochastic process. Furthermore, we've hinted towards the fact that a spectral decomposition of the conformation, the Rouse modes, allowed for an efficient description of the statistical properties of the polymer. At the end of the previous section, we pointed out that, to better exploit these results, we likely need a theoretical framework, and we anticipated that it would be based on signal processing theory.

In section 2.1, I start by recalling a few basic concepts of probability theory and give, in an intuitive fashion, elements of the mathematical theory of **stochastic processes**, a branch of probability theory.

In section 2.2, I recall the definition of the Fourier transform, in its continuous and discrete forms, introduce the Power Spectral Density, and discuss the effect of discretization of a signal on the resulting Fourier spectrum. Next, in section 2.3, the useful notion of fractional Brownian Motion, along with its spectral characterization, is introduced.

Once these well-known concepts introduced, it will be possible to connect the Rouse modes to the notion of Fourier transform, thanks to the introduction of a slightly different version of the discrete Fourier transform, the **Discrete Cosine Transform**, or DCT. In section 2.4, I will come to give an interpretation of the Rouse modes as a cosine transform of the polymer conformation. Finally, in section 2.5, building on all the concepts introduced so far, I give an original interpretation of the squared Rouse modes as a **power spectrum** of polymer equilibrium conformations.

2.1 STOCHASTIC PROCESS THEORY: DEFINITIONS

2.1.1 *Random variables*

Any measurement operated on a physical system is to some extent random. In classical physics, this perceived randomness emerges from the inevitable lack of perfect information about either the system at study, the measuring apparatus, or both. The mathematical structure that formalizes this stochasticity is called a **random variable**.

A random variable x is defined by the combination of two mathematical objects. First, the set of all possible outcomes of x, commonly noted Ω, and called the **sample space** of x.

Second, the probability to measure x in a specific interval of Ω which is called the **probability distribution function** of x, noted $P(x)$, such that

$$\mathrm{Probability}(x < \mathrm{x} < x + dx) = P(x)dx. \tag{2.1}$$

To characterize x, we can compute a number of quantities related to the shape of its probability distribution, called the moments of $P(x)$. The first moment is the **expected value** of x which is defined as the sum of all possible values of x weighted by their respective probability,

$$\mu_{\mathrm{x}} = \langle \mathrm{x} \rangle = \int_{\Omega} dx\, x\, P(x)\,. \tag{2.2}$$

The expected value of x is related to the **empirical average** of a collection of samples of x through the law of large numbers. Say we draw N samples of x, $x_1, x_2, \cdots, x_N$, the empirical average of the sample,

$$\overline{\mathrm{x}} = \frac{x_1 + x_2 + \cdots + x_N}{N} \tag{2.3}$$

converges to $\langle \mathrm{x} \rangle$ when the number of samples goes to infinity. In a real experiment, $\langle \mathrm{x} \rangle$ is impossible to compute with infinite precision, we can only approach it by computing the empirical average $\overline{\mathrm{x}}$. In statistics, $\overline{\mathrm{x}}$ is called an **estimator** for $\langle \mathrm{x} \rangle$.

The second order moment, or **variance**, is the expected value of the centered random variable associated to x, $\widetilde{\mathrm{x}} = \mathrm{x} - \langle \mathrm{x} \rangle$, squared:

$$\langle (\mathrm{x} - \langle \mathrm{x} \rangle)^2 \rangle = \int_{\Omega} dx\, (x - \langle \mathrm{x} \rangle)^2\, P(x)\,, \tag{2.4}$$

and the pth order moments is defined as the expected value of the pth power of the centered random variable $\mathrm{x} - \langle \mathrm{x} \rangle$.

As for the first one, the second order moment is related to a quantity defined on a collection of N samples, the average of $\widetilde{\mathrm{x}}^2$ or **empirical squared deviation from the mean**

$$\sigma_{\mathrm{x}}^2 = \overline{(\mathrm{x} - \overline{\mathrm{x}})^2}, \tag{2.5}$$

with $\sigma_{\mathrm{x}} = \sqrt{\sigma_{\mathrm{x}}^2}$ the standard deviation of x.

Given two random variables x and y one can compute their covariance or **correlation**[1], which quantifies the dependence of one random

---

[1] More precisely, the correlation between x and y is given by their covariance normalized by the product of their standard deviations: $\mathrm{Cov}_{\mathrm{xy}}/\sigma_{\mathrm{x}}\sigma_{\mathrm{y}}$.

variable with respect to the other:

$$\mathrm{Cov}_{\mathrm{xy}} = \langle(\mathrm{x} - \langle\mathrm{x}\rangle)(\mathrm{y} - \langle\mathrm{y}\rangle)\rangle \tag{2.6}$$

$$= \int_{\Omega_x}\int_{\Omega_y} dxdy\,(x - \langle x\rangle)(y - \langle y\rangle)\,P(x,y), \tag{2.7}$$

where $P(x,y)$ is called the **joint probability** of x and y and is defined as

$$\mathrm{Probability}(x < \mathrm{x} < x + dx \text{ and } y < \mathrm{y} < y + dy) = P(x,y)dxdy. \tag{2.8}$$

From these elementary concepts, we can now move on to define stochastic processes and investigate their properties.

2.1.2 *Stochastic Processes*

An ensemble of random variables uniquely indexed by a set of numbers J, i.e. $\{\mathrm{x}_t \mid \forall t \in J\}$ forms a function with random output and is called a **stochastic process** or **stochastic signal**. The process is said to be discrete if J is countable, and continuous if J is an interval of the real line. A typical example is a time-dependent process, in which case t represents time (and may be continuous or discrete, depending on the measure and recording method). However, the monomer number in a polymer chain can also play the role of the (discrete) variable t, that may be noted n in that case. For the sake of simplicity[2], let's consider a discrete, finite, and real-valued stochastic process of length T, with $J = \{1, 2, \dots T\}$ and

$$\mathrm{x} = \{\mathrm{x}_1, \dots \mathrm{x}_T\}. \tag{2.9}$$

The sample space of x is the union of the samples spaces of $\mathrm{x}_1, \dots \mathrm{x}_T$

$$\Omega = \Omega_1 \cup \Omega_2 \cdots \cup \Omega_T = \mathbb{R}^T. \tag{2.10}$$

In this case, a sample of x can be seen as a T dimensional vector.

We can assign a probability distribution function to the whole process, that is the joint probability

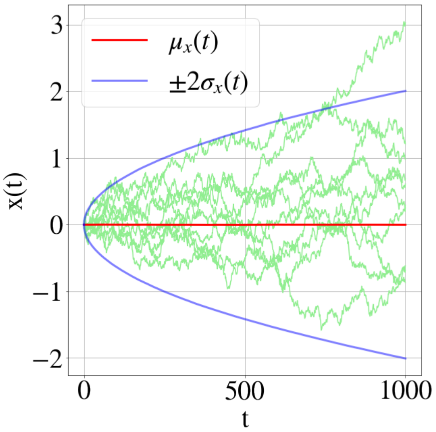
$$P(\mathrm{x}) = P(x_1, \dots x_T)\; dx_1, \dots dx_T. \tag{2.11}$$

Then, focusing on a particular t, we can define the **mean function** of x_t as the expected value of the function x evaluated at time t

$$\mu_{\mathrm{x}}(t) = \mu_{\mathrm{x}_t} = \langle\mathrm{x}_t\rangle = \int_\Omega d\mathrm{x}\, P(x_1 \dots x_T) = \int_{\Omega_t} dx_t\, P(x_t), \tag{2.12}$$

where $P(x_t) = \int dx_1 dx_2 \dots dx_{t-1} dx_{t+1} \dots dx_N P(x_1, \dots x_N)$ is the **marginal probability**[3] of x_t. The same can be done of course for the next moments. Moreover, the generalization to a particular t in a continuous process leads to analogous results for the corresponding instantaneous $\mathrm{x}(t)$.

Hence, each variable x_t (or $\mathrm{x}(t)$) can be treated as a single stochastic variable, its empirical average (or variance) on a collection of N realizations of the processes treated as an estimator of $\mu_{\mathrm{x}}(t)$ (or of the



[2] This is also always the case for digitalized signals. In any case, completely equivalent results exist for continuous signals.

[3] The marginal probability is the probability of a single event occurring, independent of other events.

variance), and so on. Nevertheless, the value of x_t is potentially dependent on the other elements of the stochastic process x, i.e. on the values of the process at other instants t'. To give a measure of this dependence, we define the **auto-correlation function** as the expected value of the product of the process evaluated at times t and t'

$$C_{\mathrm{xx}}(t,t') = \langle \mathrm{x}_t \, \mathrm{x}_{t'} \rangle, \quad \text{or} \quad \langle \mathrm{x}(t) \, \mathrm{x}(t') \rangle. \tag{2.13}$$

Note that, by definition, $C_{\mathrm{xx}}(t,t) = \sigma_{\mathrm{x}}^2$. For $t \neq t'$, vanishing $C_{\mathrm{xx}}(t,t')$ means that x_t and $\mathrm{x}_{t'}$ are statistically independent.

Random processes can be classified based on many different criteria. One of the important questions that we can ask about a random process is whether its statistical properties are time-dependent. If a stochastic process has its moment (as mean, variance, etc.) constant in time (i.e. our t parameter) it is said to be a **stationnary process**. Even more generally, a *wide sens stationary* process is a process whose auto-correlation function is transitionally invariant and whose mean function is constant, i.e.

$$\mu(t) = \mu, \tag{2.14}$$

$$C(t,t') = C(\tau = t - t') \tag{2.15}$$

where τ indicates the *delay* at which the two values of the signal are taken. The second moment is not necessarily constant but must be finite for any t.

For such processes, therefore, a single variable auto-correlation function $C(\tau)$ allows us to study the extent of interdependencies along the signal, such as hidden periodicities, persistence, and negative correlations. However, another valuable perspective is offered by the power spectral density, which provides equivalent insights to auto-correlation. Let's now delve into this aspect.

2.2 The Power Spectral Density of a Stochastic Signal

2.2.1 *Fourier Transform of Continuous and Infinite Stochastic signals*

→Integral Transforms

Any (integrable) realization $x(t)$ of a continuous and infinite random function can be seen as a vector on an infinite dimensional vector space of functions. Here $x(t)$ is called the **real space** representation of the vector x and is the projection of the signal vector on the Dirac distribution $\delta(t)$, so that

$$x(t) = \int_{-\infty}^{+\infty} x(t')\delta(t-t')dt' \tag{2.16}$$

This representation corresponds to decomposing the signal in its "single-time" components.

However, $x(t)$ can be expressed in another basis, where its properties may be more conveniently or insightfully displayed, by applying an

integral transformation to the vector. An integral transform T is defined through its action on a function f:

$$f'(u) = T[f(u)] = \int_{-\infty}^{+\infty} x(t)K(u,t)dt, \tag{2.17}$$

where the two variable function $K(u,t)$, the **kernel**, completely characterizes the transform.

→*The Fourier Transform*

The most important integral transform is surely the Fourier transform $\mathcal{F}$, defined as the integral transform with kernel

$$K(f,t) = e^{-i2\pi ft}. \tag{2.18}$$

The **Fourier transform** of a function $x(t)$ then reads:

$$X(f) = \mathcal{F}[x(t)] = \int_{-\infty}^{+\infty} x(t)e^{-i2\pi ft}dt\,. \tag{2.19}$$

The new function $X(f)$ is called the **frequency or Fourier space representation** of the function. Analyzing the expression of $X(f)$ allows one to uncover the **frequency components** of any continuous signal.

The discovery of the Fourier transform opened up new perspectives in signal processing, engineering, and, of course, physics, where it was first introduced. It can be contended that the Fourier transform stands as the most widely utilized mathematical tool for examining physical systems. A quote by Lord Kelvin (taken from [4]) summarizes well this statement:

> Fourier's theorem is not only one of the most beautiful results of modern analysis, but it may be said to furnish an indispensable instrument in the treatment of nearly every recondite question in modern physics.

2.2.2 *The Fourier Transform of Discrete and Finite Signals*

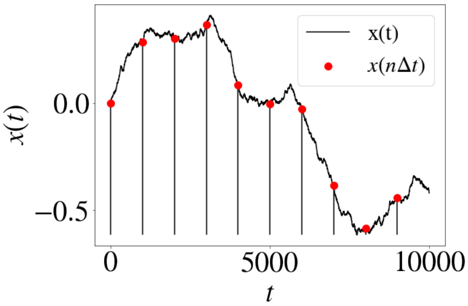
The Fourier transform applies to continuous and infinite signals. However, real signals (and polymer models as well) are always finite and discrete. For example, if we are measuring the position of a Brownian particle, we will record its position at a finite pace and for a finite time, eventually yielding a vector of positions $(x_0, \ldots x_T)$.

This measurement process introduces two time limits: the total duration T of the recording, and the sampling step Δt, that generally induce a loss of information. The total number of recordings in the sample is fixed by these two limits, as

$$N = \frac{\Delta t}{T}. \tag{2.20}$$

A discrete sampling of a continuous signal $x(t)$ is therefore written as

$$x(t) \quad\longrightarrow\quad (x(0),\, x(\Delta t),\, \ldots,\, x((N-1)\Delta t))\,. \tag{2.21}$$



---

[4] R. G. Lyons. *Understanding Digital Signal Processing*. 2010

In this case, to study the spectral properties of the signal, we need to define a discrete equivalent of the Fourier transform, that applies to finite and discrete functions. We will introduce it in a few steps, and investigate the effects of time and frequency discretization on the Fourier space representation of the signal.

→Discrete Transforms

In the discrete and finite case, instead of an integral transform, we use the term **discrete transform**. Similarly to integral transforms, discrete transforms are linear transformations that send a **discrete** vector from its original space to another vector space. Again, discrete transforms are completely determined by their kernel and their action on a vector. Let $\mathbf{x} = (x_0, \ldots, x_N)$ be a $N+1$ dimensional real vector represented in its canonical basis, and K its kernel

$$X_p = \sum_{n=0}^{N} x_n K_{n,p}, \tag{2.22}$$

The vector $\mathbf{X} = (X_0, \ldots, X_N)$ is the transformed vector by the discrete transform associated to the kernel K. In the discrete case **the kernel is an $N \times N$ transfer matrix**.

→The Discrete Time Fourier Transform

Before specifying the kernel for the discrete Fourier transform, it is instructive to take the *continuous* Fourier Transform of the *sampled* signal of equation 2.21. Formally, this is done by taking the Fourier transform of the product of $x(t)$ (continuous) and a Dirac comb:

$$X^{\mathrm{DTFT}}(f) = \mathcal{F}[\sum_{n=0}^{N-1} x(t)\delta(t - n\Delta t)] = \sum_{n=0}^{N-1} x_k e^{-i2\pi f n \Delta t} \Delta t\,. \tag{2.23}$$

The resulting function is called the **discrete time Fourier transform (DTFT)**. I stress again that the DTFT is **continuous in Fourier space** (while the signal is discrete in real space).

Let's now investigate the frequency dependence of $X^{\mathrm{DTFT}}(f)$. By introducing the **sampling frequency**,

$$f_e = \frac{1}{\Delta t}, \tag{2.24}$$

$X^{\mathrm{DTFT}}(f)$ becomes

$$X^{\mathrm{DTFT}}(f) = \sum_{n=0}^{N-1} x_n e^{-i2\pi \frac{f}{f_e} nt}. \tag{2.25}$$

Due to the periodicity of the complex exponential kernel, we see that $X^{\mathrm{DTFT}}(f)$ is now a periodic function of f with period f_e, i.e.

$$X^{\mathrm{DTFT}}(f + m f_e) = X^{\mathrm{DTFT}}(f). \tag{2.26}$$

Moreover, it can be shown[5] that **the DTFT can be expressed as an infinite summation of copies of $X(f)$**, shifted by multiples of f_e:

$$X^{\mathrm{DTFT}}(f) = \sum_{n=-\infty}^{+\infty} X(f - n f_e)\,. \tag{2.27}$$

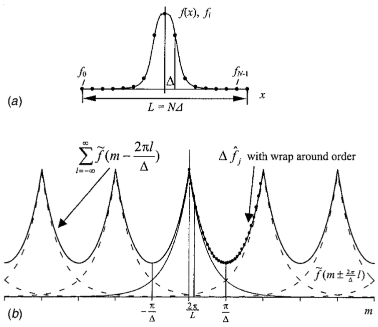


Figure 2.1: Top panel : a function $f(x)$ in real space and a discretization f_i at rate Δ Bottom panel : In dashed lines are the copies of the Fourier transform of $f(x)$, $\hat{f}(m)$. In full lines the DTFT of f_i which is equal to the pointwise sum of all the copies. Taken from [51].

[5] Here is a brief justification of this result:

The relation (2.27) derives from the fact that the Fourier transform of a Dirac comb of period T is still a Dirac comb, of period $1/T$ in the frequency space. Furthermore, the Fourier transform of the product of two functions is given by the convolution of the two Fourier transforms. The convolution of the transform of the continuous function and the frequency comb therefore gives the desired relationship.

→*The Shannon-Nyquist Theorem and The Aliasing Effect*

An important consequence of equation (2.27) is that, to ensure that repeated spectra don't overlap, the largest frequency present in the spectrum of $x(t)$, f_{max}, must not be too large compared to f_e. This limit is precised by the **Shannon-Nyquist theorem**: f_{max} must be *at most* half as a large as the sampling frequency f_e, i.e.

$$f_{max} < \frac{f_e}{2}. \tag{2.28}$$

This condition is needed in order to prevent any information loss. Indeed, if the Shannon-Nyquist condition is not satisfied, the high-frequency parts of two copies of the spectrum will meet and be summed, producing an over-representation of these frequencies. More precisely, if a signal component at frequency f is sampled at a sampling frequency $f_e < f$, the corresponding contribution will reappear, in the periodized spectrum, at the *alias* frequency $f' = (f_e - f)$. The contributions of the two frequencies f and f' - symmetric with respect to f_e - are therefore indiscernible, this leading to a loss of information about the signal.

To better understand why this happens in the real space representation, in Figure 2.2 I plot a pure sinusoidal signal of frequency f (dashed lines) sampled at a frequency lower than f: the sampled signal (dots), also perfectly matches a sinusoid of frequency $f' = (f_e - f) < f$ (full line). This effect is called **aliasing** or **spectral leakage**. The corresponding loss of information has, of course, a direct explication in real space: simply, if the signal is sampled at a too low sampling frequency $f_e = 1/\Delta t$, variations on time-scales shorter than the sampling step Δt are lost due to the sampling process, resulting in signal distortion.

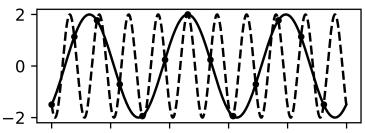


Figure 2.2: The effect of aliasing from the real space point of view: by taking a too low sampling frequency, we cannot capture the right frequency content of the underlying function.

→*The Discrete Fourier Transform*

As discussed, $X^{\mathrm{DTFT}}(f)$ is the standard Fourier transform of the sampled signal, i.e. a *continuous* function of f, which means it can't be obtained numerically because, as the signal itself, it would require infinite memory to store it. We therefore need to sample again, but this time in the continuous Fourier space of the DTFT.

A natural choice for the sampling frequency step Δf is

$$\Delta f = \frac{1}{(N-1)\Delta t} = \frac{1}{T}. \tag{2.29}$$

In this way, the N-sample in real space is transformed in a N-sample in frequency space.

For the same reasons that sampling real space periodizes the Fourier space, sampling Fourier space periodizes real space[6]. Consequently, if we sample with a frequency Δf, the copies in real space will be spaced by $1/\Delta f$. If $\Delta f = 1/T$ is chosen, then the spacing between copies in real space is T, the total length of the signal. Hence, we are ensured that copies will not overlap in real space, thus justifying the choice of the frequency step.

---

[6] In other words, inverting the discretized transform does not allow us to distinguish the initial, finite signal from its periodized version. This is related to the fact that the Fourier transform of a periodic signal is always discrete, as it is a Dirac comb, modulated by the coefficients of the corresponding Fourier series.

Such a sampling of the DFTF is called the Discrete Fourier Transform (DFT) and reads, for a generic discrete signal x = $(x_0, x_1, \ldots, x_{N-1})$,

$$\boxed{X_p^{\mathrm{DFT}} = \sum_{n=0}^{N-1} x_n \, e^{-i\frac{2\pi}{N}np},} \tag{2.30}$$

where $p = 0, \ldots N-1$.

Another way of wording this is that the DFT is the discrete transform defined through the DFT kernel,

$$\Omega_{np} = e^{-i\frac{2\pi}{N}np}, \tag{2.31}$$

or, equivalently, that the DFT can be written in matrix form as

$$\begin{pmatrix} X_0^{\mathrm{DFT}} & X_1^{\mathrm{DFT}} & \cdots & X_{N-1}^{\mathrm{DFT}} \end{pmatrix} = \mathbf{X}^{\mathrm{DFT}} = \Omega\,\mathbf{x}. \tag{2.32}$$

→ *The Fast Fourier Transform*

The computational complexity of the DFT is, a priori, $O(N^2)$, meaning it requires a number of operations proportional to N^2 to compute (each of the N modes involves a sum of N elements). However, a very efficient algorithm was discovered in 1965, by James Cooley and John Tukey called the **Fast Fourier Transform (FFT)** that cuts the complexity to $N\ln(N)$. The number of engineering situations that require the computation of a DFT is so large that the FFT is often described as one, if not the most important algorithm ever discovered.

2.2.3 *Power Spectral Density*

So far, we considered the transforms of signals which could very well be single realizations of random processes, or fully deterministic signals. If the signal is inherently random, such as in the case of noise - or for polymer conformations - each realization of the signal will have a different Fourier transform. One is then more interested in the statistical properties of the spectral content of a signal.

In many cases, deviations of the signal $x(t)$ above or below 0 are equally distributed. This property is called the **isotropy** of the real space variable. In this case, the average of the signal is zero, implying that $\langle X_p \rangle = 0$. For this reason, the spectral content of the signal is generally defined as the **square amplitude, or variance, of the Fourier space representation**, otherwise called the Power Spectral Density, that I will now introduce and that, most importantly, will finally allow us to interpret our result on the equilibrium fluctuations of the Rouse modes, equation (1.59).

→ *Signal Energy*

Generally speaking, the investigation of spectral properties in stochastic processes has its roots in communication engineering. For this reason, in the context of signal analysis, the word energy has a different meaning than in physics. The **energy E_{x} of a signal** x is defined as the integral over the whole signal of the squared magnitude of the signal. For a realization $x(t)$ of a continuous signal (on the left) - x_n in the discrete case (on the right) - of finite duration T (or N) we have

$$E_{\mathrm{x}} = \int_0^T dt\,|x(t)|^2 \qquad \text{or} \qquad E_{\mathrm{x}} = \sum_{n=0}^{N-1} |x_n|^2, \tag{2.33}$$

respectively. If we consider the energy of a centered signal averaged over several realizations

$$\langle E_x \rangle = \int_0^T dt\,\langle |x(t)|^2 \rangle, \tag{2.34}$$

we recognize the integral of the **variance** over the whole signal.

Note indeed that, most of the time, E_{x} won't have the dimensions of an energy. The term originates from the fact that most signals in engineering are a measure of voltage in a circuit. The associated (electrical) energy is

$$\frac{1}{Z}\int_0^T V^2(t)dt, \tag{2.35}$$

where Z is the impedance of the circuit, that is therefore proportional, but not identical, to E_{x}.

→ *Spectral Energy*

The previous definition of energy of course can be generalized to *infinite-length* signals, provided that the integral

$$E_{\mathrm{x}} = \int_{-\infty}^{\infty} dt\,|x(t)|^2 \tag{2.36}$$

converges: $E_{\mathrm{x}} < \infty$. Let's place ourselves within this more general framework for the moment. A signal x in real space (of any length), and its continuous or discrete transform X, are two representations of the same vector in different bases. Since the norm of a vector is invariant under a change of basis, it follows that

$$||x||^2 = ||X||^2 \tag{2.37}$$

which reads, for the continuous (on the left) and discrete (on the right) case

$$\boxed{\int_{-\infty}^{+\infty} dt\,|x(t)|^2 = \int_{-\infty}^{+\infty} |X(f)|^2\,df} \quad \text{or} \quad \boxed{\sum_{n=-\infty}^{+\infty} |x_n|^2 = \sum_{p=-\infty}^{+\infty} |X_p|^2}. \tag{2.38}$$

This important relation is called the **Parseval Theorem**.

Of course, we recognize the energy of the signal in the left-hand term of (2.38), giving alternative expressions for the signal energy:

$$E_x = \int_{-\infty}^{+\infty} |X(f)|^2\,df \qquad \text{or} \qquad E_x = \sum_{p=-\infty}^{+\infty} |X_p|^2. \tag{2.39}$$

The Parseval Theorem holds true for any integral or discrete transform.

[7] I assume here that the signal is continuous, but these results are also valid in the discrete case.

→ *The Power Spectral Density*

For a signal of infinite length[7], the energy defined by equation (2.36) may diverge.

In this case, it is commonly preferred to define an energy density, by considering truncated portions of the signal, of duration T, then dividing the corresponding energy $E_x(T)$ by the duration. This leads to the definition of the **power spectral density (PSD)**

$$\boxed{S_{\mathrm{x}}(f) \propto |X(f)|^2} \quad \text{or} \quad \boxed{S_{\mathrm{x}}(p) \propto |X_p|^2}. \tag{2.40}$$

The precise definition of the PSD for the case of continuous and infinite-length signals is briefly recalled in the following box, but these details are not relevant for the case of finite and discrete signals that will be of interest in this thesis.

Formal definition of the power spectral density

For a continuous signal of infinite length, we can consider the truncated signal

$$x_T(x) = \begin{cases} x(t) & \text{if } 0 \le t \le T \\ 0 & \text{otherwise,} \end{cases} \tag{2.41}$$

and define its Fourier transform

$$y(f,T) = \int_{-\infty}^{\infty} x_T(t) e^{-i2\pi f t} dt = \int_0^T x(t) e^{-i2\pi f t} dt \tag{2.42}$$

that always converges (provided x(t) is a continuous function).

We then define the power spectral density (PSD) of the stochastic process x as:

$$S_{\mathrm{x}}(f) = \lim_{T \to +\infty} \frac{1}{2\pi T} \, |y(f,T)|^2. \tag{2.43}$$

Following the same reasoning as above, the PSD can be seen as the spectral energy per unit time. Consequently, it describes how the power of the signal or time series is distributed over frequency.

→*PSD for a stochastic signal*

In the case of random signals, their Fourier representations are also random signals, and a single realization will generally produce a noisy PSD. Indeed, a single realization over a finite time frame cannot provide a comprehensive spectrum profile, due to the lack of statistical data. In this case, averaging over large numbers of realizations becomes crucial to improve the statistics: The average over several realizations partially "compensates" for the finite duration of the recorded signal and allows us to better approach the theoretical limit of the definition of the PSD[8].

This will be the procedure adopted in a large part of my thesis: once several realizations of a same random process are obtained, the PSD will be identified with its statistical estimator

$$\boxed{\bar{S}_{\mathrm{x}}(p) = \langle |X_p|^2 \rangle}. \tag{2.44}$$

called the **periodogram**.

[8] Several estimators of the PSD exist in signal processing. The method used here is based on a very common approach, referred to as the method of averaged periodogram. In the same way, we can obtain an estimate of the correlation function of a finite-duration signal by averaging over several realizations. This is called a **correlogram**.

Nevertheless, X_p will not be a standard (discrete) Fourier transform of the signal, as we will discover in section 2.4.

2.3 Fractional Brownian Motion and Power Law PSDs

Brownian motion describes the random movement of particles suspended in a fluid, resulting from the chaotic collisions between the particles and the surrounding molecules of the fluid. The mathematical model describing this motion, which assumes independence among particle movements, found applications well beyond the context of its discovery, in finance, chemistry, biology, etc...

→ *Random Walk and White (Gaussian) Noise*

The motion of a Brownian particle can be modeled as a 1-dimensional random walk. Such a process, that I will denote w_n, is defined as

$$w_N = \sum_{n=0}^{N} u_n \tag{2.45}$$

where the random **increments** $\mathbf{u_n}$ are uncorrelated and Gaussian, with mean $\mu_u = 0$ and variance $\sigma_u^2 = \sigma^2$, so that

$$C_{uu} = \langle u_n u_m \rangle = \sigma^2 \delta_{mn}. \tag{2.46}$$

In other terms, the increments define a **discrete white noise**: By definition, white noise is a centered, stationary stochastic process with no memory. It is fully characterized by its mean and correlation function. Thanks to the Wiener–Khinchin theorem[9], the power spectrum of white noise is easily calculated from its correlation function:

$$S_u(f) = \left| \mathcal{F}[\sigma_u^2 \, \delta(\tau)] \right|^2 = \sigma_u^2 \tag{2.47}$$

The PSD of white noise is therefore constant. Hence, by analogy with white light, its appellation.

Starting from the properties of the white noise, one can show that the correlation function of the random walk w_n is given by

$$C_{ww} = \langle w_n w_m \rangle = \sigma^2 \, \min(n, m). \tag{2.48}$$

Continuous versions $w(t)$ and $u(t)$ can also be introduced. We will find them as particular cases of the generalized signals introduced hereafter.

→*Fractional Brownian Motion and Fractional Gaussian Noise*

Despite its relevance in nature, many random phenomena cannot be modeled by Brownian motion because they experience **long range correlations**, meaning their future motion is affected by all their past positions. This leads to the introduction of a generalized model.

[9] The **Wiener-Khinchin theorem** states indeed that, for a stationary process, the Fourier transform of the correlation function is equal to the PSD:

$$S_\mathrm{x}(f) = \mathcal{F}[C_\mathrm{xx}](f).$$

The simplest model for a long-range correlated process is the **fractional Brownian motion (fBm)** $w_H(t)$ which is defined as a continuous, one-parameter stochastic process with mean function

$$\mu_{\mathrm{H}}(t) = \langle w_H(t) \rangle = 0 \tag{2.49}$$

and the following auto-correlation function:

$$C_{w_H w_H} = \langle w_H(t) w_H(t') \rangle = \frac{\sigma^2}{2}\left(|t|^{2H} + |t'|^{2H} - |t' - t|^{2H}\right). \tag{2.50}$$

As for the (discrete) random walk, the fBm process can be obtained as the integral of its increment process $u_H(t)$, also called **fractional Gaussian noise (fGn)**, and it is possible to show that the increment auto-correlation function is[10]

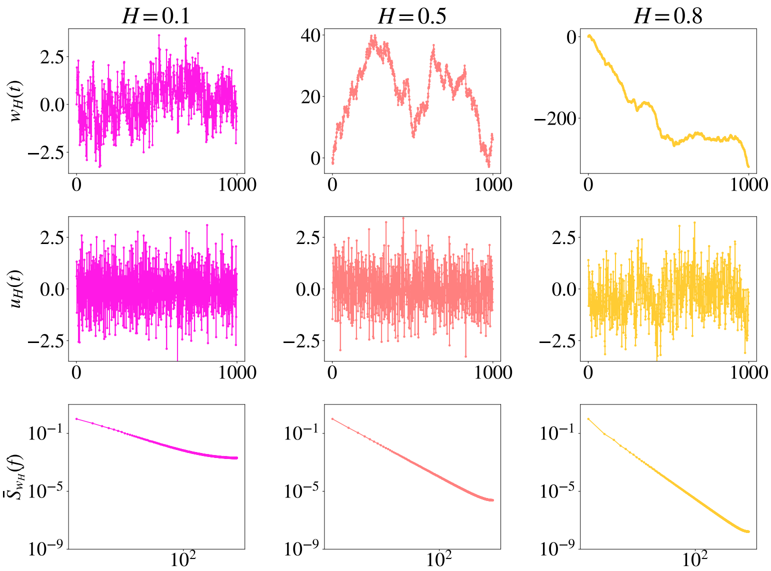


Figure 2.3: Three instances of fbm with the corresponding fgn and power spectra

$$\begin{aligned} C_{u_H u_H} &= \langle u_H(t) u_H(t') \rangle = \qquad (2.51)\\ &= 2H(2H-1)\sigma^2 |t - t'|^{2H-2} + 2H\sigma^2 |t - t'|^{2H-1} \delta(t - t'). \end{aligned}$$

The parameter H is a number in $]0,1[$ called **Hurst exponent**. It governs the correlations of the process increments. There are 3 possible correlation regimes for the increments of the process, depending on the value of H:

- $H > 1/2$ means the increments are positively correlated;
- $H < 1/2$ means the increments are anti-correlated;
- $H = 1/2$ the increments are uncorrelated.

By definition, in the **specific case** $H = 1/2$, the previous correlation functions become

$$C_{ww} = \sigma^2 \mathrm{min}(t, t') \tag{2.52}$$

---

[10] H. Qian. *Fractional Brownian Motion and Fractional Gaussian Noise*, pages 22–33. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003

and

$$C_{uu} = \sigma^2 \delta(t - t') : \tag{2.53}$$

the fBm $w_{1/2}(t)$ appears then as a continuous equivalent of the random walk (called **Wiener process**), and its increment process $u_{1/2}(t)$ defines a **continuous Gaussian white noise** with variance σ^2.

The fBm can therefore be seen as a generalization of the Brownian motion where some memory has been introduced through a choice of positively or negatively correlated steps (i.e. an increased or decreased probability that a step is in the same direction that the previous ones) and resulting, consequently, in persistent or anti-persistent motion. Figure 2.3 give some examples of fBm and of the corresponding fGn, where the effect of positive or negative correlations can be observed.

→*Power Spectral Density*

It is possible to compute the scaling of its PSD from Equation 2.43. This is done for arbitrary H in Ref.[11], and leads to the following scaling behavior:

$$\boxed{S_{w_H}(f) \sim \frac{1}{f^{2H+1}} \sim f^{\alpha}} \tag{2.54}$$

It is therefore possible to interpret the **exponent $\boldsymbol{\alpha}$** $= 2H+1$ in the power law PSD of a given signal **as an alternative signature for the correlation strength of the underlying stochastic process**:

- $-2 < \alpha < -3$ means the increments are anti-correlated;
- $-1 < \alpha < -2$ means the increments are positively correlated;
- $\alpha = -2$ the increments are uncorrelated.

This is illustrated by the third row in Figure 2.3, where the spectra corresponding to fBm of different H are shown. The slope of $S_{w_H}(f)$ in the log-log plot corresponds to α.

Note that the power law behavior of the PSD is typically associated with the self-similar or **fractal nature** of the process, as previously discussed regarding $R(s)$ in Equation 1.2.1. This behavior can be quantified by the parameter H, or equivalently, α."

This is a useful example of the complementarity between power spectral density and correlation function in the characterization of a stochastic process.

With these results in mind, let's return to the interpretation of Rouse mode amplitude fluctuations (1.59).

[11] P. Flandrin. On the spectrum of fractional Brownian motions. *IEEE Transactions on Information Theory*, 35(1):197–199, 1989

2.4 CONNECTING THE ROUSE MODES TO THE DISCRETE COSINE TRANSFORM

2.4.1 *The Rouse Modes as a Discrete Transform of a Polymer Conformation*

We are finally in a position to establish the link between chapter 1 and what we've covered so far in this chapter.

At the end of chapter 1, through the diagonalization of the Hamiltonian of the Gaussian model, we introduced a **change of basis**, defined by the **transfer matrix** Φ, and operating on the conformation of the polymer conformation $\mathbf{R} = (\vec{R}_0, \vec{R}_1, \dots \vec{R}_N)$. The elements of the transformed conformation are the well-known Rouse modes of the polymer, first presented in (1.47) and whose formula we remind below:

$$\vec{X}_p = \frac{1}{N+1} \sum_{n=0}^{N} \vec{R}_n \cos\left(\frac{p\pi}{N+1}(n+\tfrac{1}{2})\right).$$

If we isolate a single component of $\vec{X}_p$, say the x component, by projecting it on the basis vector $\vec{e}_x = (1,0,0)$, we get

$$X_{p_x} = \vec{X}_p \cdot \vec{e}_x = \frac{1}{N+1} \sum_{n=0}^{N} R_{n_x} \cos\left(\frac{p\pi}{N+1}(n+\tfrac{1}{2})\right) \tag{2.55}$$

Rouse modes as a DCT

We immediately recognize that *the three components of the Rouse modes are in fact a discrete transform of the corresponding spatial components of the conformation,* $\mathbf{R_x}$. The kernel of the corresponding discrete transform is the transfer matrix

$$\Phi_{np} = \frac{1}{(N+1)} \cos(\frac{p\pi}{N+1}(n+1/2)) \tag{2.56}$$

such that the pth mode of the new transform for a signal x_n reads:

$$X_p^{\text{new}} = \frac{1}{N+1} \sum_{n=0}^{N} x_n \cos\left(\tfrac{p\pi}{N+1}\,(n+\tfrac{1}{2})\right).$$

Now, a quite remarkable fact is that this discrete transformation is actually well known in the field of signal processing, where it's known as the **type 2 Discrete Cosine Transform** (DCT-II), or DCT for short. **The DCT is a Fourier-related transform**, in that it also projects a signal onto a set of sinusoidal functions, and each basis function corresponds to a single frequency component:

$$\boxed{X_p^{\text{DCT}} = \frac{1}{N+1} \sum_{n=0}^{N} x_n \cos\left(\tfrac{p\pi}{N+1}(n+\tfrac{1}{2})\right)}. \tag{2.57}$$

Like the DFT, the DCT is an invertible transform and therefore contains all the information about the original signal that can be recovered from the inverse DCT transform, sometimes referred to as the iDCT. This has also already been presented in the case of Rouse modes in equation (1.48), which we recall here, projected on the x axis:

$$\vec{R}_n = \vec{X}_0 + 2\sum_{p=1}^{N} \vec{X}_{p_x} \cos\left(\frac{p\pi}{N+1}(n+\tfrac{1}{2})\right).$$

It is both surprising and fortunate that the DCT should be the transformation that diagonalizes the Rouse model Hamiltonian, as it possesses specific properties that make it extremely useful in spectral analysis. Notably, its **ability to concentrate a maximum of signal variance in the smallest number of low-frequency modes**, while leaving the amplitudes of high-frequency modes close to 0. For this reason, DCT is the most widely used transform for lossy signal compression. Indeed, instead of transmitting the total N data points of a signal, one can send only the significantly non-zero modes to the receiver, which will contain an essentially unaltered version of the original signal. For this reason, the DCT is the subject of thorough research thus its properties are well known, and aggressively optimized algorithms, based on the fast Fourier transform exist and permit its efficient computation.

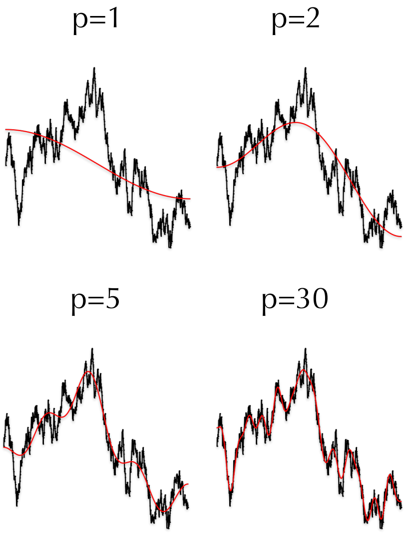


Figure 2.4: Reconstruction of a stochastic signal (1D simple random walk) of length $N = 1000$ using an increasing amount (p) of DCT modes.

→*Periodization induced by the DCT*

The DCT, just like the DFT and any discrete Fourier-related transform, induces a periodization of the signal due to the periodicity of their kernel. As stated in the previous section, the DFT implies indeed a forward copy of the signal. Let $\hat{x}_{\mathrm{DFT}}(t)$ be the original signal and $\hat{x}(t)$ the periodized signal, we have:

$$\hat{x}_{\mathrm{DFT}}(t+mT) = x(t) \quad | \quad t \in [0,T] \tag{2.58}$$

where m is an integer and T is the total duration of the signal. As most random signals don't start and end at the same point, the DFT provokes a **discontinuity** in the periodized signal (see Figure 2.5, bottom panel), which is known to reduce the rate of converges of the transform, meaning it requires more sinusoids to represent the function for a given accuracy.

However, DCT involves symmetrizing the signal before copying it. We can construct a symmetrized signal $y(t)$ from $x(t)$ as follows

$$y(t) = \begin{cases} x(t) & 0 < t < T \\ x(-t) & -T < t < 0. \end{cases} \tag{2.59}$$

$$\hat{x}_{\mathrm{DCT}}(t+mT) = y(t) \quad | \quad t \in [-T,T] \tag{2.60}$$

Both $\hat{x}_{\mathrm{DFT}}$ and $\hat{x}_{\mathrm{DCT}}$ are depicted in Figure 2.5.

Hence, by construction, the periodization induced by the DCT completely avoids the discontinuity issue. This essential difference between both transforms explains why the DCT is so popular for signal compression.

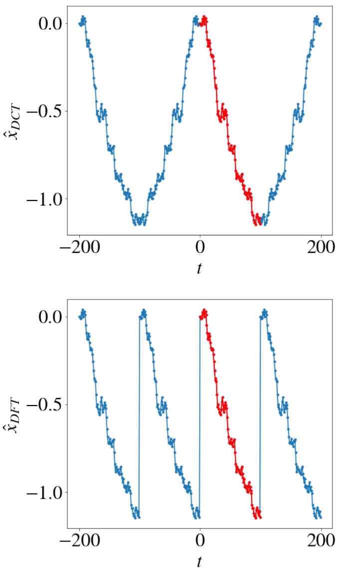


Figure 2.5: Periodization implied by the DCT (top panel) and the DFT (bottom panel) for the same stochastic signal $x(t)$ of length $T = 100$ (red).

2.5 INTERPRETING THE SQUARED ROUSE MODES AS A PSD OF THE POLYMER CONFORMATION

We can now come back to our ultimate goal in this chapter, which is the interpretation of the expression (1.59),

$$\langle X_p^2 \rangle = \frac{3}{8}\frac{k_B T}{k(N+1)}\frac{1}{\sin^2\left(\frac{p\pi}{2(N+1)}\right)},$$

for the average amplitude of the Rouse modes $\langle X_p^2 \rangle$ of the Gaussian chain.

2.5.1 *Connecting Rouse modes, DCT spectrum and random walk*

Consider again a discrete 1-dimensional random walk

$$w_N = \sum_{n=0}^{N} u_n$$

with Gaussian **increments $\mathbf{u_n}$** as previously defined (equation (2.45)).

Let's calculate the squared DCT of the signal w_N, i.e. the DCT-based spectrum of the discrete random walk[12]. Without any loss of generality, we can set $W_0^{\mathrm{DCT}} = 0$, which amounts to set the average of the signal to 0. For $p > 0$, we have by definition

$$\begin{aligned}\langle {W_p^{\mathrm{DCT}}}^2 \rangle &= \left\langle \left(\frac{1}{(N+1)} \sum_{n=0}^{N} w_n \cos\left(\tfrac{p\pi}{N+1}(n+\tfrac{1}{2})\right) \right)^2 \right\rangle \qquad (2.61)\\ &= \frac{1}{(N+1)^2} \sum_{n,m=0}^{N} \langle w_n\, w_m \rangle \cos\left(\tfrac{p\pi}{N+1}(n+\tfrac{1}{2})\right) \cos\left(\tfrac{p\pi}{N+1}(m+\tfrac{1}{2})\right).\end{aligned}$$

By using equation (2.48), we get then

$$\begin{aligned}\langle {W_p^{\mathrm{DCT}}}^2 \rangle &= \sigma^2 \frac{1}{N^2} \sum_{n,m=0}^{N} \min(n,m) \cos\left(\tfrac{p\pi}{N+1}(n+\tfrac{1}{2})\right) \cos\left(\tfrac{p\pi}{N+1}(m+\tfrac{1}{2})\right)\\ &= \sigma^2 \frac{1}{(N+1)^2} \sum_{m=0}^{N} \left[\sum_{n=0}^{m-1} n \cos\left(\tfrac{p\pi}{N+1}(n+\tfrac{1}{2})\right) \cos\left(\tfrac{p\pi}{N+1}(m+\tfrac{1}{2})\right) \right.\\ &\qquad \left. + \sum_{n=m}^{N} m \cos\left(\tfrac{p\pi}{N+1}(n+\tfrac{1}{2})\right) \cos\left(\tfrac{p\pi}{N+1}(m+\tfrac{1}{2})\right) \right]. \qquad (2.62)\end{aligned}$$

The last sum can be computed by converting to complex exponentials and using geometric series summation formulas, or by using `Mathematica`.

[12] We may note that w_n is not invariant to time translations, and the Wiener-Khinchin theorem doesn't apply here. Nevertheless, we are here interested in the operational definition of equation (2.43).

Interpretation of (1.59)

Computing the sum yields:

$$\boxed{\langle W_p^{\mathrm{DCT}^2} \rangle = \frac{\sigma^2}{8} \frac{1}{(N+1)} \frac{1}{\sin^2\left(\frac{p\pi}{2(N+1)}\right)}}. \tag{2.63}$$

The last expression allows us to finally connect the different facets of our story: what we find here is that **the DCT-based PSD of a random walk** (equation (2.63)) **is equal to the amplitude fluctuations of the Rouse modes** (equation (1.59)with

$$\sigma^2 = \frac{3k_B T}{k} = b^2 \tag{2.64}$$

where the last equality follows from equation (1.27) and shows that the average step length of the random walk w_n must indeed coincide with the average value of the bonds of the Gaussian chain for the two models to match.

In retrospect, this equality may seem almost self-evident, since the conformations of the Gaussian model are random walks. However, to my knowledge, this link has never been formally stated, let alone studied. It's striking how different the two paths to this result are.

We'll see in later chapters that this analogy provides useful insights when studying more complex polymer models. For now, let's see what else this analogy between spectral energy and Rouse modes can teach us in the simpler context we've considered so far.

To begin with, we're going to study the $p \ll N$ limit of (2.63). But before we're in a position to tackle this question, we need to go off on a little tangent by introducing a class of correlated stochastic processes, the fractional Brownian motion, a generalization of Brownian motion, and calculate their spectral properties. These results will be also useful in the next chapters.

2.5.2 *The $p \ll N$ limit and the power law prefactor*

In subsection 1.5.2, assuming $p \ll N$, we were able to expand the sine in equation (1.59) to first order in p/N, giving us the following approximation for the spectrum

$$\langle X_p^2 \rangle \approx \frac{b^2}{2\pi^2} \frac{N+1}{p^2} \sim p^\alpha \quad | \quad \alpha = -2 \tag{2.65}$$

which, in light of the previous discussion, is easy to interpret. Indeed, equation (2.65) shows that **the Rouse model spectrum is proportional to the PSD for Brownian motion**, (see equation (2.54)).

Again, the identity $b = \sigma$ has a direct interpretation ; moreover, we

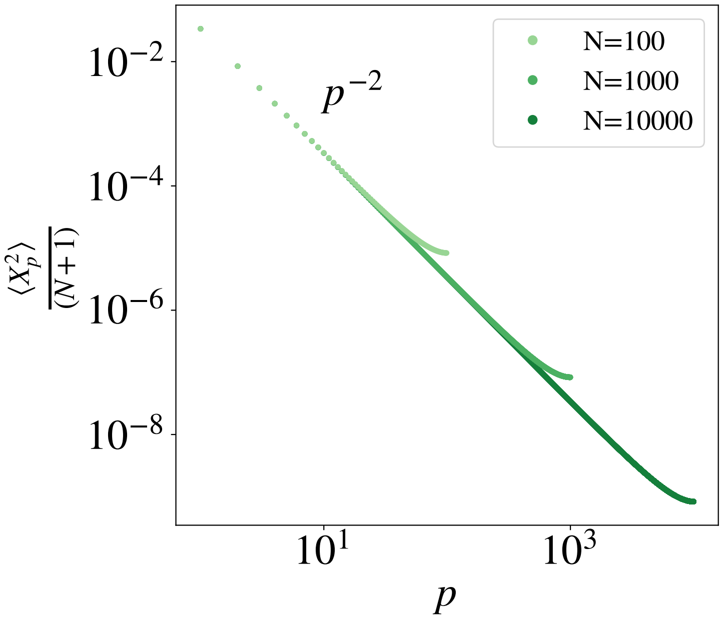


Figure 2.6: The low-p behaviors of the spectra defined by equation (2.65) for different values of N collapse on the same p^{-2} power law by taking $\langle X_p^2 \rangle/(N+1)$, according to equation (2.65).

can explain the factor $N+1$, which comes from the fact that we're comparing an energy spectrum with a power spectrum.

Remembering that lower p modes are associated with large spatial scales, the assumption $p \ll N$ can be interpreted as considering only the large-scale features of the signal. This essentially means that, looking at it from afar, the conformation of a discrete Gaussian chain is identical to the trajectory of a Brownian particle - provided that N is large enough. Again, this is not very surprising in hindsight, but while the expanded expression (2.65) is present in virtually every polymer physics book [13,14,15], this link is never discussed.

Figure 2.6 illustrates the power law behavior of (2.65) and its dependence on N.

2.5.3 *The aliasing interpretation of the large p limit*

As p approaches N, the expansion of the sine function becomes invalid and the expression (2.63) for $\langle X_p^2 \rangle$ deviates from the power law behavior. This deviation is clearly visible in Figure 2.6 for the largest values of p, and can be understood thanks to the signal-conformation analogy developed in this chapter. Indeed, aliasing offers an interpretation for this discrepancy. We can think of the Gaussian model, whose conformations are random walks, as a *discrete sampling* of continuous Brownian motion realizations. According to equation (2.54), the Brownian motion spectrum is $\sim p^{-2}$, i.e. it has non-zero contributions to all of its infinite amount of modes. It thus would need an infinite sampling rate to be accurately sampled. Any sampling will therefore induce *aliasing*: this explains the observed departure from the $1/p^2$ behavior in the high-frequency modes.

[13] A. Khokhlov. *Statistical Physics of Macromolecules.* AIP series in polymers and complex materials. AIP Press, 1994

[14] I. Teraoka. *Polymer Solutions, An Introduction To Physical Properties.* Wiley-Interscience, 2002

[15] G. Strobl. *The Physics of Polymers: Concepts for Understanding Their Structures and Behavior.* Springer Books, 1996

2.6 Conclusion / take-home messages

The main aim of this chapter was to introduce the novel concept of polymer conformation power spectral density (PSD), defined from the Rouse modes of a polymer, and to present some of its properties through the instructive example of the Gaussian chain.

To this end, based on the analogy between polymer conformations and stochastic processes, we established a previously unexplored link between the Rouse modes of a chain and a discrete cosine transform (DCT) of a signal.

By extending this, we were able to interpret the formula for the mean square of the Rouse modes as a power spectral density, allowing us to define the concept of a polymer's spectral density as:

$$\langle X_p^2 \rangle = \left\langle \left[\frac{1}{N+1} \sum_{n=0}^{N} \vec{R}_n \cos\left(\frac{p\pi}{N+1} (n + \tfrac{1}{2}) \right) \right]^2 \right\rangle .$$

which represents the average frequency contributions to chain fluctuations.

Digging deeper into this connection, we were able to interpret several properties of this PSD:

- Firstly, its asymptotic power-law dependence can be interpreted as a signature of the fractal nature of the chain. The exponent of this power law provides information on the correlations between the monomers.

- Equally important, we have seen that DCT, and therefore Rouse modes, have the property of concentrating a maximum of polymer information in a small number of low-frequency modes. This property will prove crucial for the study of the coil-globule transition in the next chapter, since, as we shall see, the collapse of the chain implies a significant modification of its first modes.

- Finally, we were able to interpret the deviation of the spectrum from this power law at high frequencies as a consequence of aliasing, hence connecting it to the discreteness of the signal.

Now that the concept of PSD has been properly defined, we can deploy it in the context of ”real polymer” models, where, as we shall see, its relevance is even greater.

3

Real Polymer Models and the Coil-Globule Transition

In the previous sections, ideal chain models were considered. Such models neglect volume interactions – i.e. all interactions between segments that are far apart along the chain but potentially close in 3D space – allowing their conformations to be mapped to the outcome of a simple random walk. In contrast, real monomers in a fluid have finite volumes that cannot overlap and, through their interactions with the solvent and themselves, can effectively attract one another. As the effective attraction increases, the polymer abruptly transitions from a decondensed coil state to a dense globular state. As we've seen in the case of ideal chains, the state of the polymer is generally identified by the scaling of its gyration radius. This means that, given a set of conformations at a fixed polymer length N, the information of its size alone isn't sufficient to characterize its state: one needs instead several data sets, at different chain lengths, to capture its scaling and deduce its state.

In this chapter, **I propose a phenomenological method, based on the spectral analysis** previously developed, to determine, from a set of conformations at a *single* polymer length, its folding state, and use it to draw the **phase diagram** of the finite size coil-globule phase transition. As we'll see in the following chapter, the spectral method developed here will be particularly well suited to analyze the chromatin conformation data from Bintu et al.[1].

In section 3.1, I start by recalling classical results concerning the **coil-globule phase transition**, including the polymer size scaling in each phase and an introduction to finite-size effects. Then, in section 3.2, I present the **interacting self-avoiding walk (iSAW)** model and a simulation scheme for generating iSAW conformations. In section 3.3 and section 3.4, I analyze the **power spectral density** for the polymers in the coil and globule phase, respectively.

[1] B. Bintu, L. J. Mateo *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, 2018, [9].

Starting from this characterization, I'll define in section 3.5 a spectral-based **order parameter** for the coil-globule phase transition, that enables me to draw the phase diagram of the transition.

Finally, in section 3.6, to prepare the data analysis to come in the next chapter, I study the consequences of **sub-sampling** the polymer on the power spectral density.

The main results presented in this chapter have been published in the paper

Assessing the polymer coil-globule state from the very first spectral modes.
by Timothy Földes, Antony Lesage and Maria Barbi
Physical Review Letters 127.27 (2021): 277801,

also available at https://hal.science/hal-03466632.

3.1 OVERVIEW OF THE COIL-GLOBULE PHASE TRANSITION

3.1.1 *Volume Interactions*

→ *Excluded volume and self-avoiding walk*

In specific conditions - typically for high temperature or high affinity between monomers and solvent molecules - the only contribution to monomer-monomer interaction is due to steric repulsion. Excluded volume due to one monomer influences the position of all other monomers, inducing long-range interactions along the chain and restricting the accessible conformations to those where the chain doesn't overlap with itself. The resulting conformations are no longer simple random walks but instead as **self-avoiding walks (SAW)**, i.e. three-dimensional walks that do not visit the same point more than once. The polymer thus occupies more space resulting in swollen, elongated conformations called **coils**.

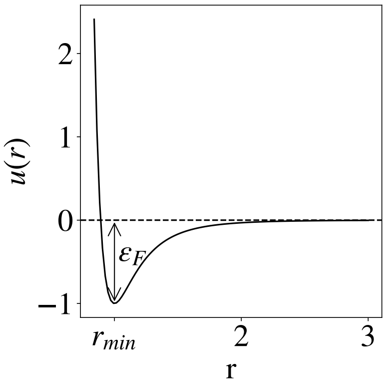


Figure 3.1: The Lennard-Jones potential $u(r) = 4\epsilon_F((\sigma/r)^{12} - (\sigma/r)^6)$ is typically chosen to model volume interaction potential. The parameter ϵ_F controls the depth of the potential well, while σ specifies the distance at which the potential is zero, often referred to as the 'size of the particle'. The position of the minimum is fixed by the value of σ: $r_0 = \frac{\sigma}{2^{1/6}}$

→ *Monomer attraction and effective potential*

On the other hand, when strong effective attractive interactions occur - either through van der Waals forces or a poor affinity with the solvent molecules - monomers are forced together, spatially confining the polymer and producing curled-up conformations called **globules**.

A single volume interaction potential $u(r)$ can be chosen to model both attractive and repulsive interactions between two non-bonded monomers located at a distance r from each other. Thanks to universality, in the thermodynamic limit, its specific functional form does not influence the scaling properties of the resulting models, provided it has a form close to that schematically represented in Figure 3.1. At short distances, it must be strongly positive - typically diverging as r tends towards zero - so to account for excluded volume. It must then become attractive at intermediate distances, where it must display a potential well, and tend towards zero at long distances as the attraction vanishes. The distance at which $u(r) = 0$, noted σ, controls the excluded volume or size of a monomer and the depth of the potential well, while the amplitude parameter[2] ϵ_F controls the intensity of attractive interactions. In Figure 3.1, the most popular example of such a potential, the Lennard-Jones (LJ) potential is shown.

The state of a solvated polymer depends only on the *relative* strength of the inter-monomer forces and the thermal energy provided by the fluid, accounted for by the **energy parameter** $\epsilon = \epsilon_F/k_BT$, that I will use all along this thesis.

3.1.2 *Gyration radius scaling across the coil-globule phase transition*

The usual observable used to qualify the polymer's state is its average **gyration radius**, defined in chapter 2 as the root-mean-square distance of the monomers to the center of mass of the polymer, that gives a measure of the polymer's size (see equation (1.14)). Its scaling with respect to the number of monomers N, for $N \gg 1$, follows a power law:

$$\langle R_g^2(N)\rangle \sim N^{2\psi} \tag{3.1}$$

where the value of the exponent $\psi = 1/d$, with d the fractal dimension of the polymer, discriminates between different polymer states.

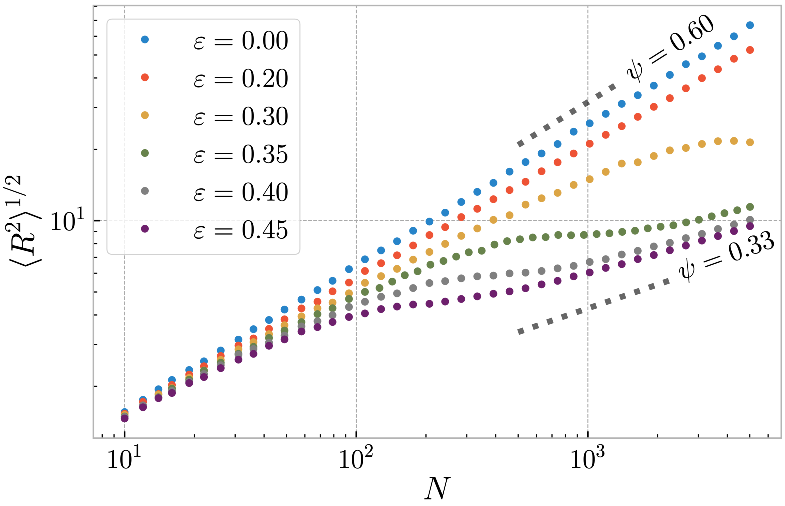


Figure 3.2: The gyration radius is plotted in log-log scale versus the degree of polymerization N for different values of ϵ. The figure is taken from [47].

In the coil state, the swollen polymer's gyration radius scales as N^ν, where $\nu \approx 0.588$ is the Flory exponent; in the globular state, the collapsed polymer acts like a space-filling curve and $R_g \sim N^{1/3}$. A graphical representation of the scaling of the R_g function of N for different values of ϵ is depicted in figure 3.2.

---

[2] Normally noted ϵ. The subscript $_F$ is here added to differentiate from the adimensional energy parameter $\epsilon = \epsilon_F/k_BT$.

In the scaling limit, $N \to \infty$, the polymer undergoes a genuine phase transition, called polymer collapse or **coil-globule phase transition**, at a particular value of the energy parameter called the **θ-energy** and noted ϵ_θ hereafter. For $\epsilon < \epsilon_\theta$ steric interactions prevail and swollen conformations are favored, while for $\epsilon > \epsilon_\theta$ attractive interactions dominate and globular states are observed. At $\epsilon = \epsilon_\theta$, attractive and repulsive effects compensate, and the polymer behaves like an ideal chain, mimicking the ideal random walk scaling $R_g \sim N^{1/2}$. Note that the value of ϵ_θ *isn't universal* as it depends on the choice of $u(r)$ and the bonding potential.

Size scaling across the coil-globule phase transition

Summary of conditions and scaling exponents expected for the three typical polymer folding states.

	Coil	θ-conditions	Globule
Condition:	$\epsilon < \epsilon_\theta$	ϵ_θ	$\epsilon_\theta < \epsilon$
scaling exponent:	3/5	1/2	1/3

3.1.3 *Finite size effects*

In the case of finite-sized polymers [3], the sharp collapse at ϵ_θ is replaced by a more complex, continuous transition of **finite width** $\Delta\epsilon$. At finite polymer length, for ϵ slightly below the $\epsilon \in [\epsilon_\theta, \epsilon_\theta + \Delta\epsilon]$ the attractive forces still aren't enough to fully collapse the polymer into a proper globule. Instead, the polymer enters the **crossover regime** or **transition phase** in which its state is ill-defined, and still poorly understood. When approaching $\epsilon_\theta + \Delta\epsilon$, the polymer goes from a denser ideal chain at energies slightly under ϵ_θ to a state with a globular core and decreasing density towards its edges[4]. Due to the cooperative nature of the transition, **the width of the transition $\Delta\epsilon$ increases as the number of monomers decreases**, meaning it requires more energy to fully collapse a smaller chain.

3.2 MONTE CARLO SIMULATION OF ON-LATTICE COLLAPSING POLYMERS

Apart from the thermodynamic scaling of polymer size, derived by Flory and reported in Figure 3.1.2, theoretical results concerning the coil-globule transition are notoriously difficult to obtain, requiring complex field-theoretic tools and rather crude approximations[5]. Consequently, major recent theoretical advances have been made possible by the deployment of computer simulations, which have repeatedly invalidated field-theoretic predictions[6,7].

For this reason, to study the spectral behavior of collapsing chains, we will not attempt to derive exact results, but instead use extensive in-silico experiments, followed by a phase of extensive statistical analysis of the results.

---

[3] J. des Cloizeaux and G. Jannink. *Polymers in solution. Their modelling and structure.* Oxford university press, New York, 1991, [18].

[4] A. Lesage, V. Dahirel *et al.* Polymer coil–globule phase transition is a universal folding principle of drosophila epigenetic domains. *Epigenetics & Chromatin*, 12(1), May 2019, [47].

[5] J. des Cloizeaux and G. Jannink. *Polymers in solution. Their modelling and structure.* Oxford university press, New York, 1991, [18].

[6] P. Grassberger and R. Hegger. Simulations of three-dimensional θ polymers. *The Journal of Chemical Physics*, 102(17):6881–6899, 1995, [30].

[7] P. Grassberger. Pruned-enriched rosenbluth method: Simulations of θ polymers of chain length up to 1 000 000. *Phys. Rev. E*, 56:3682–3693, Sep 1997, [29].

In the following section, I begin by introducing and justifying the use of lattice-based polymer models, and present a model relevant to the study of chain collapse. I then briefly describe the program I used for simulating this model.

3.2.1 *On-lattice models and the Interacting Self Avoiding Walk Model*

Thanks to the universal nature of the coil-globule phase transition, many of its properties can be studied using any type of bonding and volume interaction potentials, provided that some essential properties are ensured. In fact, we can even restrict the space in which the polymer evolves to a **discrete space**, i.e. the position of each monomer $\vec{R}_n$, instead of tacking values in $\mathbb{R}^3$, can live on some discrete lattice without altering universal features of the transition[8].

Of course, the most common choice is the cubic lattice $a\mathbb{Z}^3$ where a is the distance between two neighboring lattice sites, but other more elaborate lattices can be chosen for algorithmic optimization.

Simply sampling N points in $a\mathbb{Z}^3$ will obviously not generate a proper polymer conformation. The minimum requirement is to add the chain connectivity. The simplest way is to fix the bond length to the lattice spacing a simply meaning that two bonded monomers must be placed on neighboring sites. If only this requirement is enforced, configurations of this model will be simple random walks on the cubic lattice and thus will belong to the ideal universality class along with the FJC and the Gaussian Chain. To generate real polymers, we need to add more restrictions to model volume interactions.

The excluded volume between beads is modeled by enforcing that each lattice site be occupied by at most one monomer. Technically, this can implemented by assigning infinite energy to any conformation containing overlapping monomers. Without any other restriction, **the resulting conformations are SAWs and model polymers in the high-temperature phase**, where attractive interactions are completely neglected.

To model the monomer-monomer attraction, however, we can add an energetic proclivity for 3D **nearest neighbor (NN) contacts** by subtracting a fixed amount of energy ϵ_F for each NN contact. For a self-avoiding conformation, the energy of a given conformation writes:

$$E = -n_c \epsilon_F \tag{3.2}$$

Where n_c is the number of monomers that are one lattice edge apart in the conformation. Of course, the parameter ϵ_F controls the intensity of the attractive interaction, identically to its homonym defined earlier. The resulting model is called the **interacting Self Avoiding Walk (iSAW)**[9].

All in all, in the canonical ensemble, the probability of a conforma-

---

[8] The specific choice of the lattice again won't change the scaling behavior of the resulting models but only the non-universal features, typically the value of ϵ_θ.

[9] M. C. Tesi, E. J. Janse van Rensburg *et al.* Monte carlo study of the interacting self-avoiding walk model in three dimensions. *Journal of Statistical Physics*, 1996, [77].

tion $\mathbf{R}$ in the iSAW model is given by

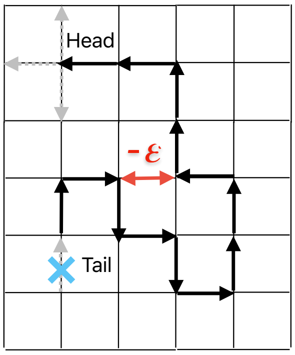
$$P(\mathbf{R}) = \begin{cases} \frac{1}{Z} e^{\frac{n_c \epsilon_F}{k_B T}} = \frac{1}{Z} e^{n_c \epsilon} & \text{if connectivity isn't broken and beads} \\ & \text{don't overlap} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where Z is the canonical partition function of the model and we have defined the energy parameter $\epsilon = \epsilon_F / k_B T$. As we'll see further on, the iSAW model presents all the features of the coil-globule phase transition and is extremely efficient to simulate.

3.2.2 *Monte Carlo Simulations of the On-lattice Interacting Self Avoiding Walk Model*

For this part, I used a high-performance, optimized Monte Carlo simulation code written in Rust by Antony Lesage during his thesis. He was kind enough to provide it to me and explain how it works and how to run it.

To draw conformations from Equation 3.3, we employ the Monte Carlo sampling method with the Metropolis-Hastings acceptance ratio and reptation moves. The idea of this move is to delete the tail of the polymer and add, in a randomly chosen direction, a new monomer at the head of the polymer. The motion of such a polymer, in Monte Carlo time, resembles that of a snake, thus the street name of this algorithm: *the slithering snake*. The algorithm goes as follows



1. We start from an initial SAW configuration.
2. Compute the initial conformation's energy E_{old} by counting NN contact.
3. Pick randomly an orientation for the polymer (head or tail).
4. Delete the last tail monomer.
5. Pick randomly a position for the new head. Reject conformation if self-avoidance isn't respected.
6. Compute the new conformation's energy E_{new}
7. Accept the new conformation with the Metropolis-Hastings acceptance ratio
$$\Delta = \max\left(1,\, e^{-(E_{\text{new}} - E_{\text{old}})/k_B T}\right)$$

It is, of course, highly inefficient to record the conformation every time a move is accepted. It is preferable to record a set of conformations that are completely decorrelated from one another, i.e. to wait for a decorrelation time τ between two conformation captures. The decorrelation time has been conjectured to be $\tau = 1/2N^2$ and does, indeed, allow decorrelated conformations to be obtained over the ϵ range explored.

In the next section, I analyze, to the best of my knowledge for the first time, the equilibrium amplitude of the Rouse modes of simulated polymers in a wide range of N and ϵ and give an interpretation of the result using the analogy between the Rouse mode and the power spectral density developed in chapter 2.

3.3 THE POWER SPECTRUM OF A COIL POLYMER

In this section, I analyze the power spectrum of *purely repulsive* polymers, the prototypical coil polymers, and show that it follows a characteristic power-law dependence in p. By comparison with attractive chains, this result will allow formal characterization of polymers in the *coil* state.

3.3.1 *Self Avoiding Walk Scaling*

In the large N limit, when the attraction ϵ_F is much lower than $k_B T$, the repulsive interaction between monomers dominates and the chain behaves as if no attractive interaction was present. The fundamental difference between this chain and the Gaussian chain is that the volume occupied by a monomer is not accessible to the rest of the chain. Conformations of these polymers can be thought of as random walks where steps are chosen to avoid earlier visited positions, a.k.a self-avoiding walks (SAW). Owing to the universality of the macro features of the coils[10], one can study the simplest SAW, the cubic lattice SAW, and expect its statistics to be identical to other coils.

Since each monomer has to avoid all other monomers, the position of one part of the chain depends on the position of the rest of the chain (in spatial dimension smaller[11] than 4). Consequently, the self-avoidance property induces **positive long-range correlations** between steps that affect its large-scale statistics. The internal end-to-end vector amplitude, for a large enough segment s, exhibits a power law behavior

$$R^2(s) \sim s^{2\nu} \tag{3.4}$$

where $\nu = 0.588 > 0.5$ in 3D[12]. Consequently, the effect of self-avoidance is to **swell** the physical dimension of the coil with respect to an ideal chain (see Figure 3.3).

Importantly, the power-law behavior of the internal end-to-end distance (3.4) indicates that, as for ideal polymers, the conformations of the coil are **statistical fractals**. The different critical exponent indicates their place within different universality classes.

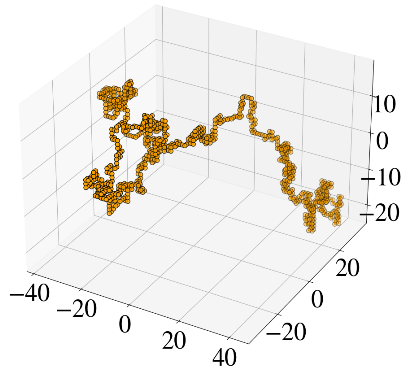


Figure 3.3: A single conformation of a coil

3.3.2 *Extending the Rouse modes analysis to a self-avoiding polymer model*

In principle, it is possible to build a model for a self-avoiding polymer by adding volume exclusion to the bead-spring model[13] through the introduction of a purely repulsive potential force between all monomers (the same of Figure 3.1, shifted by ϵ_F and truncated at the bottom of its minimum, for example). However, doing so adds non-linear terms to

---

[10] A. Khokhlov. *Statistical Physics of Macromolecules.* AIP series in polymers and complex materials. AIP Press, 1994, [43].

[11] In spatial dimension larger than 3, $R^2(s) \sim s$, like the simple RW.

[12] J. Dayantis and J.-F. m. c. Palierne. Scaling exponents of the self-avoiding-walk-problem in three dimensions. *Phys. Rev. B*, 49:3217–3225, Feb 1994, [16].

[13] Or Gaussian chain model.

the Hamiltonian. The Rouse modes, in principle, no longer diagonalize the Hamiltonian and, consequently, the equation

$$\langle \vec{X}_p(0).\vec{X}_q(t)\rangle \sim \frac{N}{p^2} e^{-\frac{t}{\tau_p}} \delta_{pq} \tag{3.5}$$

no longer holds and no analytical result for their equilibrium amplitude is known.

However, D. Panja and G. Barkema, in the context of polymer dynamics, proposed an approximate formula [14] for the correlation function of the Rouse long-wavelength modes of the self-avoiding polymer. For $N \gg 1$, they make the ansatz that

$$\langle \vec{X}_p(0).\vec{X}_q(t)\rangle \approx A_1 \frac{N^{1+2\nu}}{p^{1+2\nu}} e^{-A_2 \frac{N^{1+2\nu}}{p^{1+2\nu}} t} \delta_{pq}\,, \tag{3.6}$$

where ν is the Flory exponent $\nu \simeq 0.588$ and A_1 and A_2 are two constants. This assumes that different modes are uncorrelated. Although this isn't exact, it holds up to a very good approximation and allows the analytical computation of dynamical observables.

When evaluated in $p = q$ and $t = 0$, Equation (3.6) gives an approximate formula for the PSD of the self-avoiding polymer:

$$\langle \vec{X}_p^2\rangle \approx A_1 \frac{N^{2\nu}}{p^{1+2\nu}} \propto \frac{1}{p^{1+2\nu}} = \frac{1}{p^{\alpha}}, \tag{3.7}$$

with, therefore, exponent $\alpha = 1 + 2\nu \simeq 2.176$.

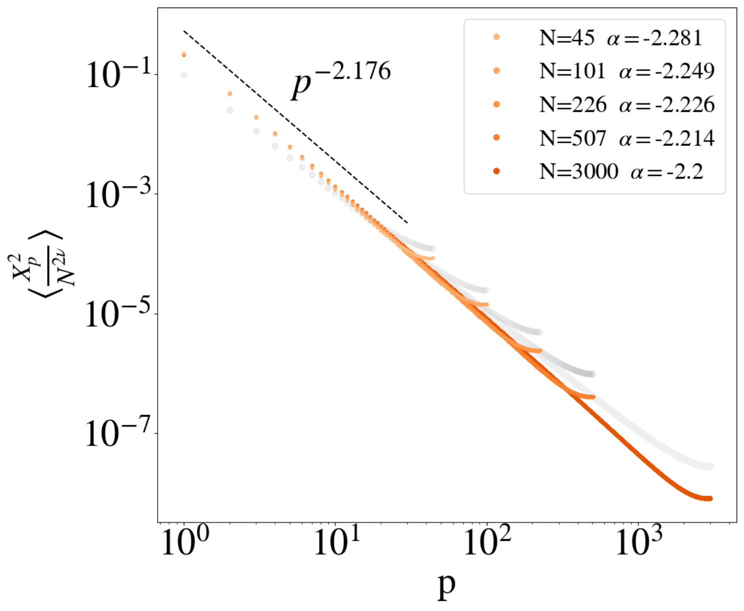


Figure 3.4: In orange the PSD, $\langle X_p^2\rangle$, for simulated coil polymers of different length N is depicted. In light grey, the PSD of a Gaussian chain polymer is shown for comparison.

In Figure 3.4 I plot the PSD of self-avoiding on-lattice polymers and find that it indeed obeys the power law predicted in [15] (Equation (3.7)). The discrepancy in the exponent for small values of N, seems to be primarily a finite size effect, since it disappears as N increases.[16]

[14] D. Panja and G. T. Barkema. Rouse modes of self-avoiding flexible polymers. *The Journal of Chemical Physics*, 131(15):154903, Oct. 2009, [58].

[15] D. Panja and G. T. Barkema. Rouse modes of self-avoiding flexible polymers. *The Journal of Chemical Physics*, 131(15):154903, Oct. 2009, [58].

[16] By inspection of Figure 3.4 we may note, furthermore, that for large p, similarly to the PSD of the ideal polymer, the power spectrum is found to deviate from its power law behavior. As discussed in subsection 2.5.3 this discrepancy is due to the discrete nature of the polymer model.

go to ToC

PSD scaling for the coil polymer

We conclude that the **characterizing feature of the PSD** of the coil state is a power law decay with exponent

$$\langle X_p^2 \rangle \sim p^{\alpha} \quad | \quad \alpha = 2\nu + 1 \approx -2.17 \tag{3.8}$$

Thus, the coil state is easily identifiable from the PSD of a given polymer: Given a set of equilibrium conformations, we can calculate the PSD of the polymer and, if we find a power law decay, the polymer is an ideal coil if $\alpha = 2$, or an expanded coil if $\alpha = 2\nu + 1$.

3.3.3 *Interpretation of the coil PSD*

The Equation (3.7) is given in Ref. [17] on a heuristic basis, without specific rationale. In light of the spectral interpretation of the Rouse modes, I propose an original interpretation of such a power law decay in the $\langle X_p^2 \rangle$, as follows.

The fractal nature of the SAW means that its statistical properties are scale invariant[18,19,20]. Hence, we expect a power law PSD, intuitively related to the presence of long-range correlations induced by steric interaction. Moreover, the same power law decay was observed in the PSD of **fractional Brownian motion** (fBm) introduced in section 2.3. As we have seen, the Hurst exponent H quantifies the correlations of the increments. The comparison with the size scaling $\langle R^2 \rangle \sim N^{2\nu}$ would lead to the ansatz $H = \nu$ which appoints the PSD found by (3.7).

To recapitulate, for ideal chains we found that the projection of the conformations on a particular axis were simple random walks yielding $\alpha = 2$ as a power law exponent. In the case of coils, the excluded volume of monomers induces long-range correlations along the polymer and the projection of the conformation on an axis is now a random walk with **positively correlated steps** of parameter $H = \nu$, yielding $\alpha = 1 + 2\nu$.

3.4 THE POWER SPECTRUM OF GLOBULE POLYMERS

In the scaling limit $N \to \infty$, at $\epsilon > \epsilon_\theta$, the attractive interactions among the monomers are strong enough to collapse the polymer into a **homogeneous, roughly spherical shape** of densely packed monomers called the **equilibrium globule**. Due to its homogeneous density, this globular state is commonly characterized by its size scaling

$$\langle R_g^2 \rangle \sim N^{2\alpha} \sim N^{2/3} \tag{3.9}$$

the exponent $\alpha = 1/3$ being typical of the scaling of a space-filling, 3D solid object. However, the gyration radius clearly provides no

---

[17] D. Panja and G. T. Barkema. Rouse modes of self-avoiding flexible polymers. *The Journal of Chemical Physics*, 131(15):154903, Oct. 2009, [58].

[18] A. Y. Grosberg and A. R. Khokhlov. *Statistical Physics of Macromolecules*. AIP series in polymers and complex materials. AIP Press, 1994, [31].

[19] , [?].

[20] L. Peliti. *Random walks with memory*, page 73. Elsevier, Amsterdam: North Holland, 1986, [60].

information about the spatial configuration of the polymer. To remedy this, we need to describe the trajectory of the chain inside the globule.

3.4.1 *The Internal Globule Structure*

The usual description of a collapsed polymer is that of a linear chain compressed within a spherical volume of radius $\sim R_g$. Over small length scales, the polymer behaves like an **ideal chain**, until it reaches the volume boundary, where it is reflected and starts an **independent** random walk[21]. The globule can thus be thought of as a dense liquid of independent ideal chains, otherwise called a **polymer melt**. The ideal behavior of the chain on small length scales is a feature of polymer melts[22] which will be discussed in the next chapter when we turn to the dynamic aspects of polymer physics. From this analogy, we can deduce its internal end-to-end distance function

$$R^2(s) \sim \begin{cases} s & \text{for } s < s^* \\ R_g^2 & \text{for } s > s^*, \end{cases} \tag{3.10}$$

with the same critical number s^* related to the globule size. To determine s^* we can reason as follows. For sub-chains that fit into the dimension of the globule, R_g, the scaling is that of an ideal chain, while longer chains have their size limited by R_g. Hence, we can deduce that

$$R^2(s^*) = s^* = N^{2/3}. \tag{3.11}$$

It's important to note that despite the power-law dependence of the gyration radius, the globule is **not** a fractal object, since its scaling behavior changes for longer distances along the chain. The size of the compact ball formed by the stacking of all its monomers is, indeed, an intrinsic length scale of the system.

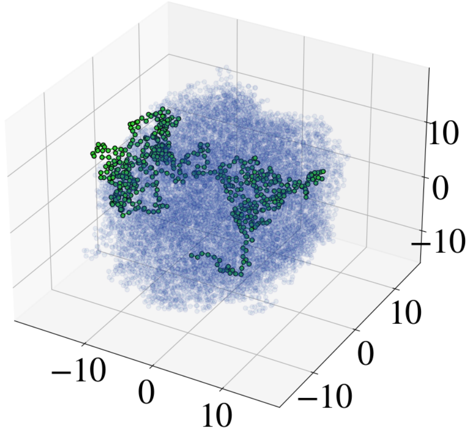


Figure 3.5: In blue a globule conformation of length $N = 12000$ and a subchain of length $g* = 12000^{2/3} \approx 624$, in green. The statistics of the subchain are expected to be Gaussian, according to standard polymer theory.

3.4.2 *PSD of globules*

The PSD of an on-lattice polymer at $N = 3000$ and $\epsilon = 0.6 \gg \epsilon_\theta$, which is a typical globular polymer, is plotted in log/log scale in Figure 3.6.

We observe that the PSD has the following structure:

- In the first few low frenquency modes, the PSD is almost **constant**, $\alpha_\mathrm{l} \approx 0$;
- In the high frequency modes $p > 30$ the PSD is found to retrieve a $\alpha_\mathrm{h} \approx 2$ power law decay, characteristic of the **ideal state**;
- A curved region links both regimes, around a critical mode number p^*, that is close to $p = 10$ in the case of Figure 3.6.
- And again we observe aliasing at the highest frequencies in the spectrum.

The PSD of the globule conformations is thus **characterized by a constant spectrum for small values of p and a power law**

---

[21] A. Khokhlov. *Statistical Physics of Macromolecules*. AIP series in polymers and complex materials. AIP Press, 1994

[22] H.-P. Hsu and K. Kremer. Static and dynamic properties of large polymer melts in equilibrium. *The Journal of Chemical Physics*, 144(15):154907, 04 2016

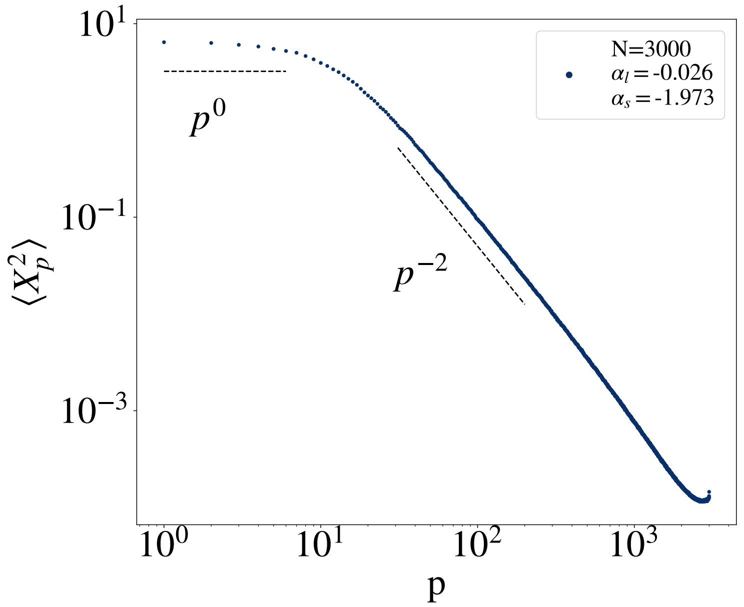


Figure 3.6: Power spectrum of an interacting polymer with interaction parameter $\epsilon = 0.5$ and $N = 3000$. The dashed lines correspond to p^0 and p^{-2}respectively. α_l and α_s are the slopes in the log-log plot between the modes 1-2 and 100-200 respectively

decay with $\alpha = 2$ **for large values of** p. In the following paragraph, I attempt to explain its shape by using the fruitful analogy between polymer conformations and stochastic processes.

Given a discrete signal $\mathbf{x} = (x_1...x_N)$ and its discrete cosine transform $\mathbf{X} = (X_1...X_N)$, **truncating the DCT at** X_n, $0 < n < N$, is equivalent to **re-sampling** $\mathbf{x}$ at a sampling rate of $\Delta = n/N$ or, equivalently, keeping only every m sample in $\mathbf{x}$ such that $N = nm$. In other words, the DCT of the re-sampled signal is equal to the truncated DCT. We notice that the re-sampling procedure is analogous to the **coarse-graining** procedure as explained in section 1.3, where, by regrouping several monomers of the freely jointed chain, I derived a statistically equivalent model, the bead spring model.

If we apply the coarse-graining procedure to the globule by truncating at the p^*th mode, where the **flat plateau** in the spectrum starts, the result is a coarse-grained polymer of n "super-monomers" or **blobs**, consisting each of m original monomers, such that $N = mn$.

Now, as we have seen, the power spectrum of such coarse-grained polymer is approximately constant, i.e. equivalent to the PSD of **white noise**, meaning that the positions of the blobs are **uncorrelated**. Thus, at large scales, the polymer can be seen as several uncorrelated blobs of m monomers.

Besides, the information about the small-scale structure of the globule is contained in the higher frequencies of the PSD where it retrieves its $\alpha = 2$ power law decay, meaning that inside the blobs, the polymer essentially behaves like an **ideal chain**. All in all, as expected, the globular state can indeed be seen as a "liquid" of uncorrelated large blobs composed of ideal polymers.

3.4.3 *Link between the critical mode p^* and the corresponding length size s^*, and conclusions*

Of course, the smallest number of monomers m^* in a blob such that the blobs are decorrelated is exactly the number of monomers between two reflections, i.e. s^*. Consequently, the value of p^* at which the plateau begins should be given by

$$\frac{N}{p^*} \sim s^* \sim N^{2/3}$$

$$p^* \sim N^{1/3} \tag{3.12}$$

This scaling law is checked in Figure 3.7 by plotting the PSD for different-sized globules function of p/p^*, which aligns the fracture points of the curves.

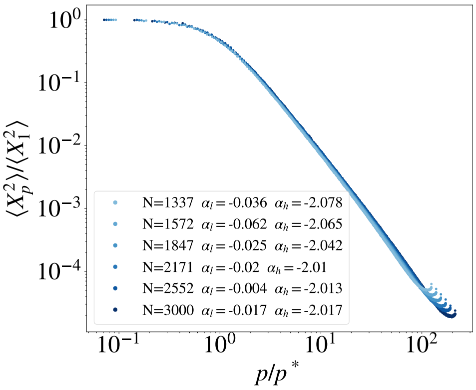


Figure 3.7: Globule PSD for several values of N function of p/p^*. The curves are seen to collapse, confirming the scaling for p^*.

PSD scaling for the globule polymer

We found that the PSD of globules is characterized by a segmented shape:

1. in the low frequency modes, $p < p^*$, it assumes a **constant spectrum**

$$\langle X_p^2\rangle \sim p^\alpha \quad | \quad \alpha = 0 \tag{3.13}$$

2. an a power law with exponent $\alpha = 2$ in the high frequency modes.

We related the PSD to the structure of the globule conformations: at small length scales, the globule behaves like a simple random walk, while at large scales, monomers are completely uncorrelated.

Given a set of conformations of an interacting polymer, since coils and globules have identical spectrums at high frequencies, to characterize its state, one can look at the low-frequency modes alone. If $\alpha \approx 0$ we are in the globule state, if $\alpha \approx 2\nu + 1$ we are in the coil state. This is a useful result since it allows to **identify the state of a polymer of given N and ϵ from a set of conformations**, i.e. without having access to the scaling of its gyration radius.

In the following section, I demonstrate the value of this result by using it to draw a phase diagram of the finite-size coil-globule phase transition.

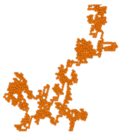
3.5 PHASE DIAGRAM OF THE COIL-GLOBULE PHASE TRANSITION

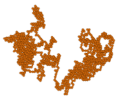
A **phase diagram (PD)** is a graphical representation of the state of a system. Each axis represents a physical parameter of the system, and at each point of the graph is associated the state of the system at those particular values of the physical parameters. The state of the system is characterized by an **order parameter**. In the thermodynamic limit, the PD is split in distinct **phases** where the system is found in a specific state defined by the value of the order parameter, separated by **critical curves** along which the system undergoes a **phase transition**.

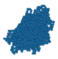
In the case of the **coil-globule** phase transition, in the thermodynamic limit, the only physical parameter is the dimensionless energy parameter $\epsilon = \epsilon_F / k_B T$. Thus, the PD is one dimensional, split by a critical point, the θ-point, that we note ϵ_θ. Determining the latter fully establishes the PD. The order parameter of the system is generally chosen as R_g [47], or a function of it, and its scaling defines the state of the polymer, as seen in table 3.1.2. An operational way of defining ϵ_θ is as the temperature (or energy) at which the repulsive and attractive forces exactly compensate. At this point, since the phase transition is of **second order**, the specific heat of the system diverges[8, 10mm].

In the finite N case, two physical parameters determine the state of the system, ϵ and N; thus the PD is two-dimensional. At finite size, the polymer goes from one state to the other smoothly, and both the sharpness and the position of the transition depend on N. The system only performs a true phase transition in the thermodynamic limit, in finite-size systems, the very concept of a transition line is ill-defined. Nevertheless, certain statistical observables show particular behavior at specific points in the transition region, allowing for the definition of **several** transition lines. Inspired by the system's critical behavior in the $N \to \infty$ limit, we notice that, at fixed N, the specific heat per monomer (the variance of the energy of a typical monomer) hits a maximum in the transition region, whose coordinates $(N, \epsilon_\theta^{c_p}(N))$ can be defined as a transition line. Similarly, the transition line can be defined as the coordinates $(N, \epsilon_\theta^{RW}(N))$ at which the polymer is in the ideal, RW-like state. These transition lines are distinct, but all converge in the thermodynamic limit to ϵ_θ.

In the following section, I propose an original way of determining the PD of the coil-globule phase transition, based on the *spectral* characterization of the coil and the globule state developed in the preceding sections, which not only gives the critical curve, but also gives an idea of how far the system is from the phase transition, thus an idea of the **transition width**.







3.5.1 *The PSD Throughout the Collapse and the order parameter α*

→*Evolution of the PSD throughout the collapse*

In Figure 3.8, the PSD of a polymer of length $N = 1137$ and interaction parameter ϵ ranging from 0 to 0.6 is shown. We recognize immediately the characteristic power-law shape of the (orange) coil PSD, and the (blue) segmented globule PSD. The black curves correspond to values of ϵ for which the polymer is undergoing the coil-globule phase transition. A global inspection of these curves allows for a few remarks.

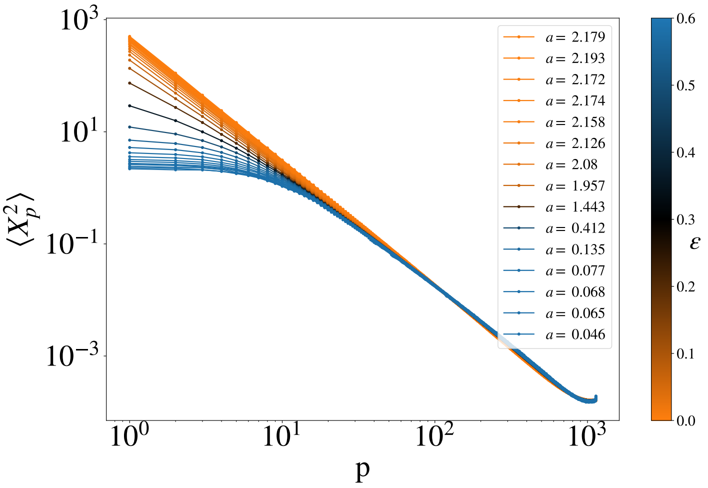


Figure 3.8: Averaged PSD calculated over an ensemble of 16384 on-lattice MC simulated configuration and over the three spatial directions. Standard deviations are smaller than the symbol size. The parameter ϵ varies from 0 to 0.6 (see color bar).

First, all the curves essentially merge in the high-frequency part of the spectrum, $p > 30$, which means that the local structure of the chain hardly changes at all during the transition. As mentioned above, on closer inspection, the slope of the high-frequency modes is in reality slightly lower in the globules, $\alpha = -2$ versus -2.17 in the coil, while the amplitude remains essentially the same.

However, the low-frequency modes change radically before and after the transition: they are greatly suppressed in the globule phase.

Now let's look closer at the behavior of the low-frequency modes during the collapse. The collapse can be decomposed into 3 stages, depicted in Figure 3.9:

- Let's start from the low-interaction or high-temperature phase, $0 < \epsilon < 0.27$ (left panel). At $\epsilon = 0$, the PSD exhibits a perfect power law with exponent $\alpha = -2.17$ as expected. As epsilon increases, the PSD retains its power-law form, but the value of α decreases slightly until it reaches the value of -2 at $\epsilon = 0.27$. This value of α is indicative of **ideal chain conformations**: hence, at $\epsilon = \epsilon_\theta = 0.27$, the polymer is in the **θ-conditions** (as usually defined from the conformation point of view).

- For $0.27 < \epsilon < 0.40$, the polymer enters the cross-over phase, during which conformational changes take place, resulting in the PSD changing dramatically. From the ideal-state power law behavior, a bend appears in the low-frequency modes whose curvature increases with ϵ, until the plateau in the low-frequency modes appears and the PSD settles in the segmented shape characteristic of the globule phase.

- At $\epsilon > 0.40$, the globule is formed and becomes increasingly compact as ϵ increases, leaving the overall shape of the PSD unchanged, with only the location of the plateau gradually decreasing.

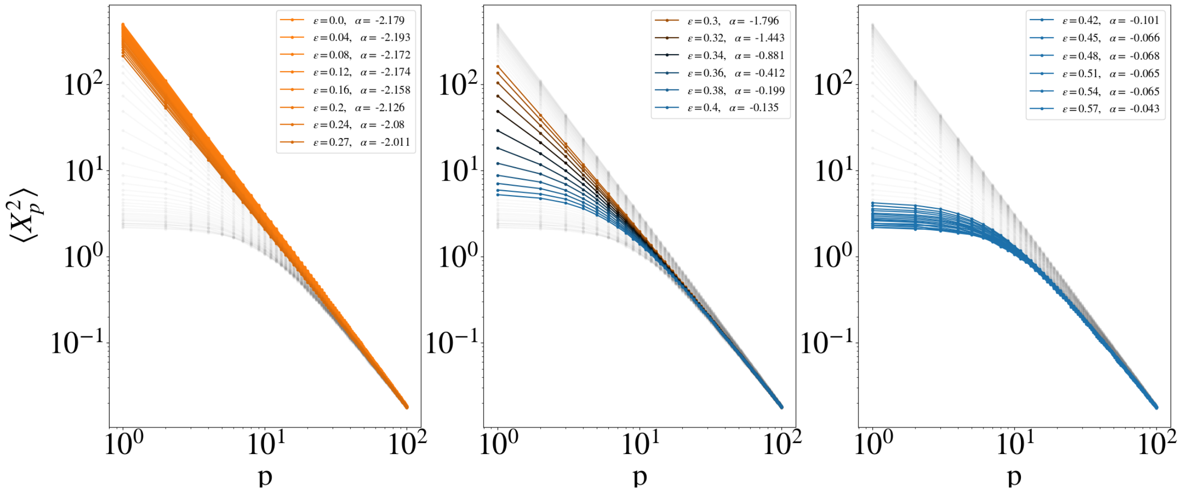


Figure 3.9: The same PSD of Figure 3.8 where the different phases of the transition are evidenced.

→*The Order Parameter* α

Clearly, the conformational changes occurring during the phase transition are concentrated in the low-frequency modes. Inspired by this fact, let's define a statistical observable, based on the low-frequency modes, that captures the change in spectral properties of the polymer.

Definition of the order parameter alpha

Let's consider the **log-log slope in the first two modes**

$$\boxed{\alpha_N(\epsilon) = \frac{\ln\langle X_2^2\rangle - \ln\langle X_1^2\rangle}{\ln 2 - \ln 1} = \frac{1}{\ln 2}\ \ln\frac{\langle X_2^2\rangle}{\langle X_1^2\rangle}.} \tag{3.14}$$

The value of $\alpha_N(\epsilon)$ for the spectrums of Figure 3.8 are plotted in Figure 3.10. As the polymer collapses, **$\boldsymbol{\alpha_N(\epsilon)}$ continuously evolves from $\boldsymbol{-(2\nu+1)\approx 2.176}$ to 0**. Due to the universal nature of the spatial fluctuation regimes in both the coil and the globule phase, α possesses the property that it has a well-defined value in both phases, making it a good choice for an **order parameter** for the transition.

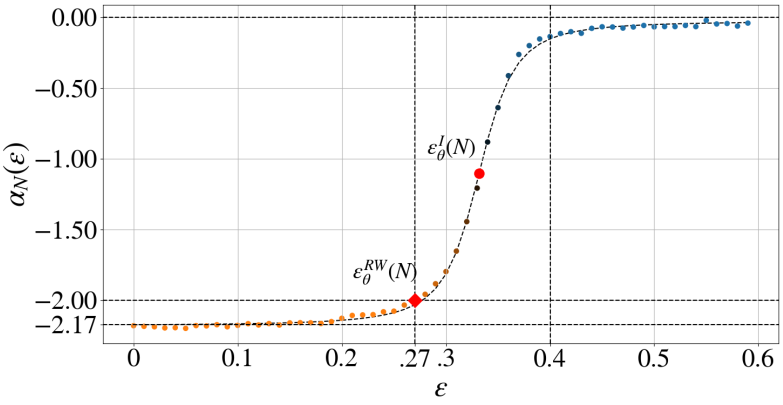


Figure 3.10: The slope $\alpha_N(\epsilon)$ of the log-log representation of the PSD, estimated from the first 2 modes, as a function of ϵ.The red dots indicate the two critical points as discussed in the text.

From the curve $\alpha_N(\epsilon)$, we can a priori define a critical point $\epsilon_\theta(N)$ which will constitute a point on the critical line of the transition. As the spectral point of view has never been used before to study this phase transition, we need to define the criterion for identifying the critical point ourselves, which we'll do in the next section.

3.5.2 *Critical Points at fixed N*

In this section, we use the spectral properties of the collapsing chain to define three statistical quantities that will identify a critical point along an iso-N line in the (N, ϵ) phase diagram. The first two can be defined from the curve $\alpha_N(\epsilon)$ and the third will be defined from the fluctuations of $\alpha_N(\epsilon)$.

→RW Critical Point

As mentioned earlier, we expect the polymer to exhibit ideal chain behavior at the so-called **Flory critical point**. Thus we define a critical point $\epsilon_\theta^{RW}(N)$ as the value of ϵ at which the order parameter α is equal to -2, i.e.

$$\boxed{\alpha_N(\epsilon_\theta^{RW}(N)) = -2} \tag{3.15}$$

In Figure 3.10, this corresponds to the intersection of the α curve with the horizontal line $y = -2$.

→Sigmoid Inflection Point

When the temperature is lowered below $\epsilon_\theta^{RW}(N)$, the polymer enters the cross-over phase. From this energy value, the rate of increase of the order parameter, or **susceptibility** of the order parameter w.r.t ϵ, defined as

$$\frac{\partial \alpha_N}{\partial \epsilon}(\epsilon), \tag{3.16}$$

rapidly increases until it hits a maximum at the mid-point of the curve, and starts decreasing again, until it goes back to 0 in the globule phase.

we can define a critical point as the value of ϵ at which the max in susceptibility is reached, i.e:

$$\boxed{\epsilon^I_\theta(N) = \text{argmax}\left(\frac{\partial \alpha_N}{\partial \epsilon}(\epsilon)\right)}. \tag{3.17}$$

To compute the value of $\epsilon^I_\theta(N)$ for a specific value of N we fit the $\alpha_N(\epsilon)$ curve with the following generic 4-parameter sigmoid function:

$$S(x) = D + \frac{A(x-B)}{\sqrt{1 + C(x-B)^2}} \tag{3.18}$$

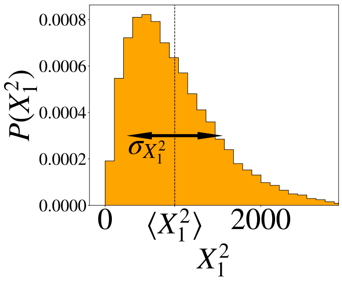
whose **inflection point**, given by the parameter B, offers a simple procedure to determine $\epsilon^I_\theta(N)$ from the order parameter curve at fixed N.

→*Relative Fluctuations in the Modes and Associated Critical Point*

The last transition line we define is related to the fluctuations of the order parameter. Let's start by defining the **relative fluctuations** c_v, otherwise called **coefficient of variation** of mode p defined as

$$CV[X_p^2] = \frac{\sigma_{X_p^2}}{\langle X_p^2 \rangle} \tag{3.19}$$

where $\sigma_{X_p^2}$ is the standard deviation of the random variable X_p^2 (see the side figure for a reminder of statistical definitions). The curve for $CV[X_p^2](\epsilon)$ for the first 9 modes is drawn in Figure 3.11.



We notice a peak in the first modes, that attenuates rapidly with p. We define our third critical line as the ϵ value at which this peak is reached. In practice, since the effect is strongest in the first mode, we'll use the relative fluctuations of X_1. Thus, we define :

$$\boxed{\epsilon^F_\theta(N) = \text{argmax}\left(CV[X_1^2](\epsilon)\right).} \tag{3.20}$$

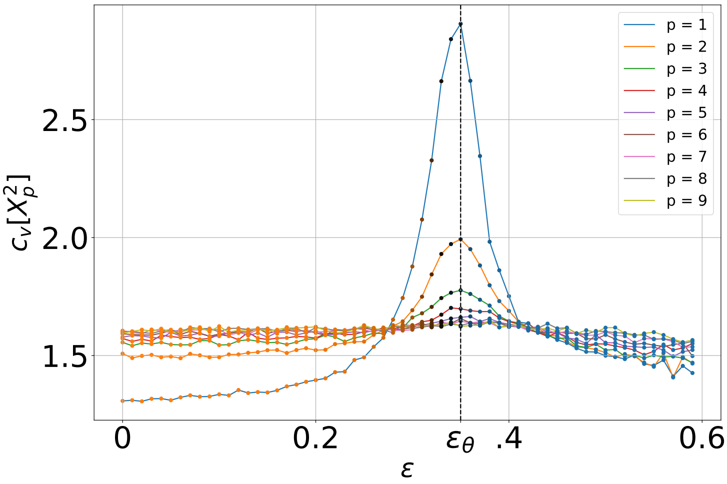


Figure 3.11: C_v curves for the first 9 Rouse modes of a polymer of size $N = 3000$. The color of the dots is based on the order parameter α and indicates the folding state of the polymer, orange for coils, blue for globules, and black for transition

In the following section, we compute all three critical lines by computing the value of all three critical energies for different values of N and plot the phase diagram of the finite-size coil-globule phase transition.

3.5.3 *Critical lines*

To compute the aforementioned critical lines, I have simulated 16000 conformations of polymers with sizes ranging from $N = 101$ to $N = 3000$ and interaction parameter ϵ from 0 to 1, then computed the value of α and CV for each simulation.

The plots of $\alpha_N(\epsilon)$ are shown in Figure 3.12. For each polymer length, by fitting the sigmoid (3.18) to $\alpha_N(\epsilon)$ we could extract the critical curves $\epsilon_\theta^I(N)$ and $\epsilon_\theta^{RW}(N)$, shown in Figure 3.12, lower plots. The fluctuations of $\alpha_N(\epsilon)$, as defined by CV in equation (3.19), are shown in Figure 3.13. By computing the max of the CV curves we extracted $\epsilon_\theta^{RW}(F)$ (Figure 3.13, bottom).

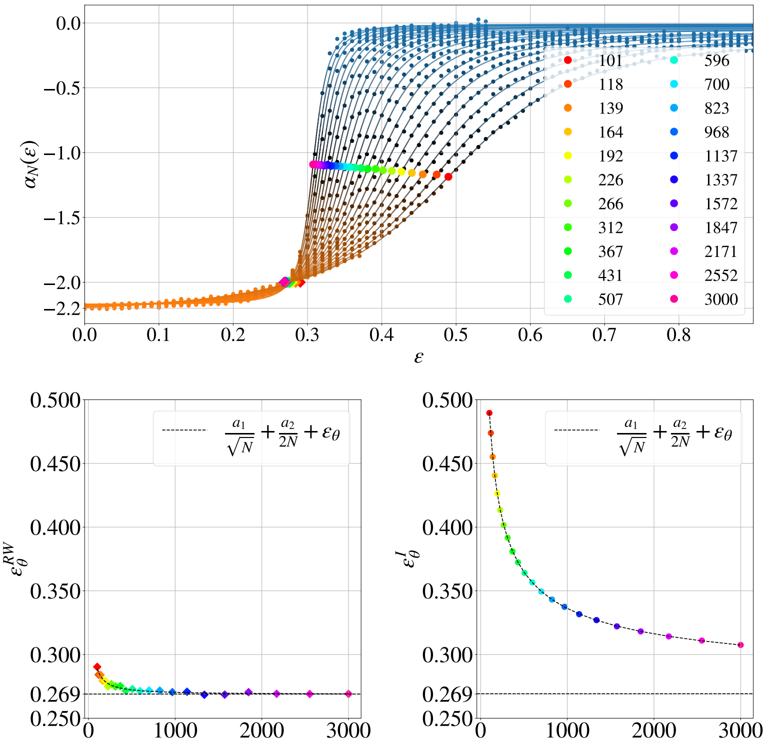


Figure 3.12: Top panel: Log-log plot slope $\alpha_N(\varepsilon)$ estimated on the first 2 modes of Power Spectral density (dots) for all polymers sizes N from 101 to 3000 (right to left). For each N, $\alpha_N(\varepsilon)$ is fitted by a sigmoid (3.18) (lines). Colored circles are inflection points, defining $\varepsilon^I(N)$.
Bottom left: the Random Walk critical line ϵ_{RW} as a function of N, fitted by (3.22) (dashed curve)
Bottom right: The sigmoid inflection point critical line as a function of N, fitted by (3.22) (dashed curve).
On both bottom plots, the horizontal dashed line corresponds to the fitted asymptotic value ϵ_θ.

No theoretical prediction is available for the critical lines based on spectral quantities, but it is reasonable to suppose that they follow a similar scaling as their classically defined counterparts - the divergence of the specific heat and the random walk scaling of the gyration radius - which follow the same scaling. Different predictions are available for this scaling: the standard mean-field prediction provided by Flory theory is

$$\epsilon_\theta(N) = \frac{a}{\sqrt{N}} + \epsilon_\theta \tag{3.21}$$

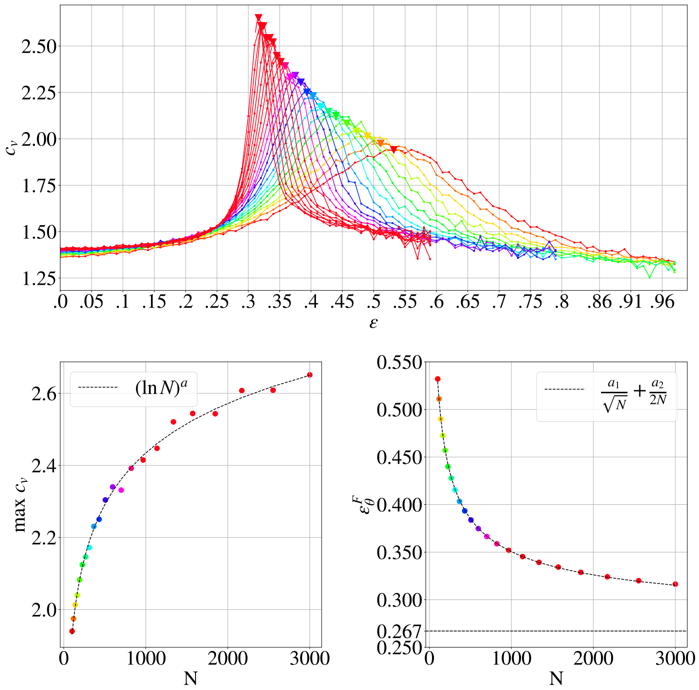


Figure 3.13: Top panel : C_v curves of the first mode for polymers of size N from 101 to 3000. The color scheme matches that of Figure 3.12.
Bottom Left: the colored points are the value of maxc_v as a function of the polymer size N. The dashed line is fit with the logarithmic function $b \log(N^a)$.
Bottom right: Value of the fluctuations maxima critical points ϵ_F function of N fitted by (3.22).

where ϵ_θ is the $N \to \infty$ transition point. Otherwise, [23] showed by on-lattice simulations that the boundary between the two phases is best fitted by a function of the form

$$\epsilon_\theta(N) = \frac{a_1}{\sqrt{N}} + \frac{a_2}{N} + \epsilon_\theta. \tag{3.22}$$

Both functions were excellent fits, but, as for [24], equation (3.22) fits our data slightly better and was consequently employed for our analysis.

Critical Lines Fitting and Asymptotic Value of the Critical Energy ϵ_θ

Fitting all three spectral-based curves previously obtained by equation (3.22) provides the asymptotic values ϵ_θ and the prefactors a_1 and a_2 given in the following table

Data	a_1	a_2	ϵ_θ
$\epsilon^I(N)$	2.1 ± 0.04	3.3 ± 0.9	0.2694 ± 0.0005
$\epsilon^F(N)$	2.5 ± 0.09	2.7 ± 1.8	0.270 ± 0.002
$\epsilon^{RW}(N)$	0.0 ± 0.1	4.6 ± 1.8	0.269 ± 0.003

As expected, all three critical lines converge to the same asymptotic value. Moreover, its value

$$\boxed{\epsilon_\theta \approx 0.269} \tag{3.23}$$

is in perfect agreement with previous numerical estimates

[23] T. Vogel, M. Bachmann *et al.* Freezing and collapse of flexible polymers on regular lattices in three dimensions. *Physical Review E*, 76(6), Dec. 2007, [81].

[24] T. Vogel, M. Bachmann *et al.* Freezing and collapse of flexible polymers on regular lattices in three dimensions. *Physical Review E*, 76(6), Dec. 2007, [81].

Specific heat divergence contradicts mean-field prediction

As a final remark, to further reinforce the relevance of our analysis, we verify our results against the field-theoretic prediction for the divergence of the specific heat c_p, given in Ref. [20]:

$$c_p \sim \ln(N)^{3/11}. \tag{3.24}$$

The specific heat is the variance of the energy, which is related to the spatial fluctuations of the polymer conformations. Thus, it is expected that the divergence in c_p should follow the same law as the mode relative fluctuations CV.

According to Ref.[30], a logarithmic divergence is indeed observed but the exponent value is off the mark. We come to the same conclusion and measure an exponent of $a \approx 0.56$.

3.5.4 *Phase Diagram*

Collecting all the information gathered in the previous sections, in Figure 3.14 we draw the **phase diagram for the finite size coil-globule phase transition**. The color of the rectangle at position (N, ϵ) represents the value of α for this pair of parameters. A diverging color map ranging from orange for $\alpha = -2.17$, to blue at $\alpha = 0$, and passing through white at the mid-value $\alpha = -1.1$, was chosen. The blue, green and black full lines represent $\epsilon^{RW}(N)$, $\epsilon^{I}(N)$ and $\epsilon^{F}(N)$ respectively, and the red dashed line represents the transition line of Vogel[25], based on the divergence of c_p.

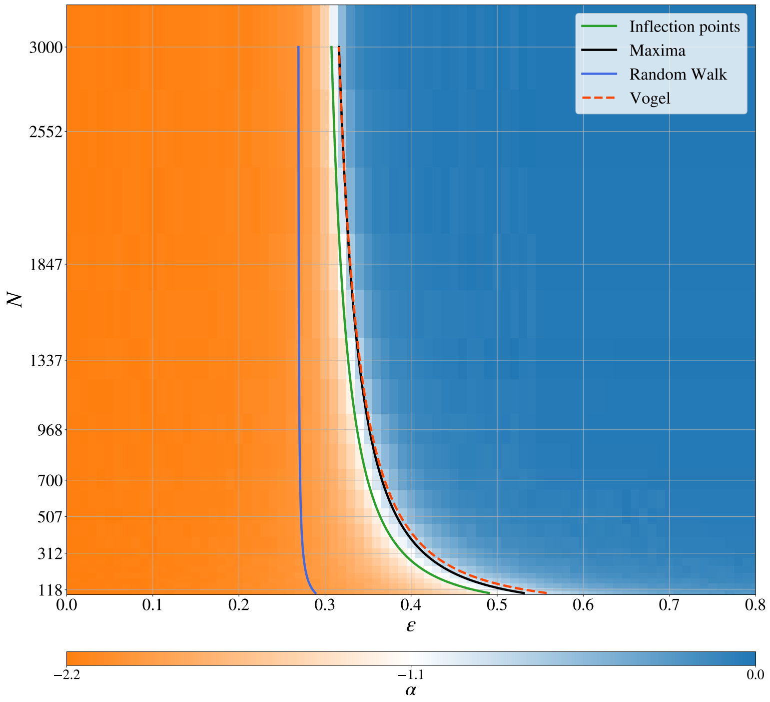


Figure 3.14: Phase Diagram for the finite size coil-globule phase transition.

[25] T. Vogel, M. Bachmann *et al.* Freezing and collapse of flexible polymers on regular lattices in three dimensions. *Physical Review E*, 76(6), Dec. 2007, [81].

Figure 3.14 summarizes nicely the thermodynamic properties of the transition partly exposed in the previous sections. Firstly, the two phases and the transition zone are clearly visible. The collective nature of the transition, namely that smaller chains require more energy to collapse, is illustrated by the ”L” shape of the transition zone. The width of the transition can also be easily gauged: we see that the white zone broadens for small N and gets more narrow as N goes to infinity, which is coherent with the $1/\sqrt{N}$ scaling predicted by mean field for the transition width[26].

We can also see that in finite-size, the Flory critical point ϵ_θ^{RW}, at which the chain adopts ideal conformations, converges much faster to ϵ_θ than the other critical lines. Consequently, it is actually relatively far from the collapse of the chain itself (inflection points) and from the point where divergent fluctuations occur (maxima), both of which are close together. The phase diagram suggests instead that the Flory temperature is the boundary between the coil phase and the transition phase. The existence of two distinct critical lines has already been pointed out by des Cloizeaux [18].

Highly fluctuating critical conformations

Hereby confirm that the conformation of a finite-size polymer at **the coil-globule transition cannot be described as a pure random walk, but rather to a complex, highly fluctuating state**, as suggested by its PSD and the divergence of CV for the low-frequency modes. To the best of my knowledge, the PSD and mode fluctuations presented in the previous section are the first descriptions of the structure of these critical conformations.

Furthermore, as we'll see in section 4.2, where we will investigate in more detail these conformations, their fluctuating nature will reveal crucial in modeling human chromatin organization.

Assessing the folding state from the value of α

That such a precise representation of the phase diagram can be drawn from the value of α indicates that the value of α is directly related to the folding state of the polymer, independently of its size or interaction parameter. In other words, α seems to be a good indicator of the level of “globulness” of the polymer. This feature allows us to **unambiguously identify the folding state of a polymer** without having to reconstruct its scaling law, as required by the usual definition of different phases.

In the next section, we'll see how this property opens up experimental perspectives for identifying the folding state of chromatin in fluorescence microscopy experiments.

---

[26] J. des Cloizeaux and G. Jannink. *Polymers in solution. Their modelling and structure.* Oxford university press, New York, 1991, [18].

3.6 ASSESSING THE COIL-GLOBULE STATE FROM A DECIMATED POLYMER

In the previous sections, we've established that, thanks to the universal values of the order parameter α, it allows assessing the state of a single polymer from a sampling of its conformations. Moreover, since this order parameter can be calculated on the basis of the first Rouse modes only, it opens important perspectives on the possibility of *experimentally* assessing the state of a polymer from a very reduced and rather accessible information.

Indeed, in order to access to the first M modes, indeed, it is in principle sufficient to record configurations of M *distinguishable* monomers, *equally spaced*[27] along the polymer and *covering the whole chain*, either by following their dynamics or by averaging over a collection of identical polymers to ensure enough statistics. This property was already introduced in section 3.4, where we showed that a subsampling of a globule conformation was, in a certain limit, equivalent to a white noise signal.

In addition, the design of multicolor fluorescent imaging of DNA sites with high spatial and temporal resolution now makes it possible to record the trajectories and relative positions of multiple loci simultaneously [28,29,30]. If the DCT-transform of a set of M-point configurations contains the equivalent of the first modes for the whole chain as we predict, this yields a new experimental approach for the determination of the state of single polymers and biopolymers.

I tested this idea on **decimated polymer chains**, i.e. reduced signals where the positions $\vec{r}_n$ of **only M equally spaced monomers** are retained, which amounts to operate a **decimation of order $m = N/M$**, i.e. keeping only every mth monomer in the chain. We decimated down to $M = 3$, an extreme condition of particular interest from the experimental point of view. Figure 3.15-a shows examples of PSD obtained for different choices of M. Extreme decimation ($M < 10$) alters the slope in the first modes (although not drastically). As a consequence, the asymptotic values of the sigmoid $\alpha_{N,M}(\epsilon)$ corresponding to "pure" coil (α_C) and globule (α_G) states vary (while, interestingly, the inflection point α_I seems to stay roughly constant), so that it is necessary to provide reference values for these limit slopes as a function of M.

These values are given in Figure 3.15-b.Despite this effect, the asymptotic values for coils and globules remain well apart down to $M = 3$ (Figure 3.15-b), even for relatively low statistical sampling as shown by the shaded areas, giving a measure of the expected confidence interval on α for different sample sizes.

Once these M-dependent limit values obtained, the order parameter **$\alpha_{N,M}(\epsilon)$ can be normalized** as

$$\tilde{\alpha}_N(\epsilon) = \frac{2\alpha_{N,M}(\epsilon) - \big(\alpha_G(M) + \alpha_C(M)\big)}{\big(\alpha_G(M) - \alpha_C(M)\big)} \tag{3.25}$$

so to span from -1 to 1. In this way, equivalent sigmoids $\tilde{\alpha}_N(\epsilon)$ are

---

[27] The requirement for uniform spacing in, in fact, not strictly necessary. If the position of the chosen monomers is known, one can use this information when projecting on the modes, and get an exact result. In the following, however, I'll limit myself to this simplest case.

[28] H. Chen, M. Levo *et al.* Dynamic interplay between enhancer–promoter topology and gene activity. *Nature Genetics*, 50(9):1296–1303, July 2018, [12].

[29] H. Sato, S. Das *et al.* Imaging of DNA and RNA in living eukaryotic cells to reveal spatiotemporal dynamics of gene expression. *Annual Review of Biochemistry*, 89(1):159–187, June 2020, [66].

[30] J. K. Eykelenboom and T. U. Tanaka. Zooming in on chromosome dynamics. *Cell Cycle*, 19(12):1422–1432, May 2020, [23].

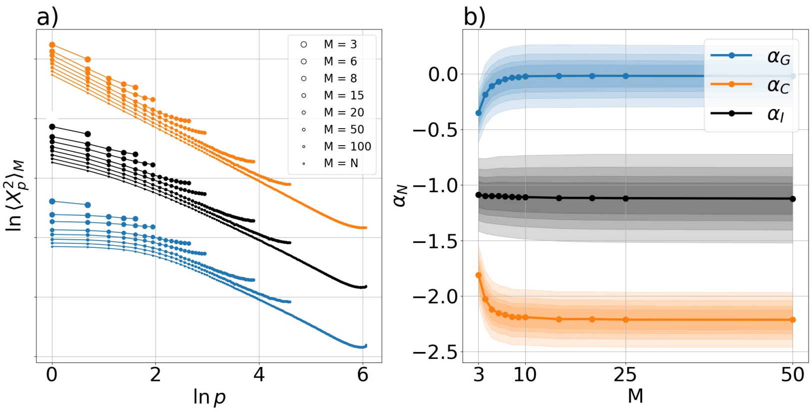


Figure 3.15: **(a)** Power Spectral Densities obtained by successive decimation of a $N = 431$ chain for $\varepsilon = 0.0$ (orange, coils), $\varepsilon = 0.32$ (black) and $\varepsilon = 0.46$ (blue, globules). Decimation goes from $M = N$ to $M = 3$ (longer to shorter curves). **(b)** The asymptotic and critical values of $\alpha_{N,M}$ obtained from the sigmoid fit of the decimated spectra, as a function of M (dots and lines): $\alpha_C(M) = \alpha_{N,M}(0)$ for coils, in orange; $\alpha_I(M) = \alpha_{N,M}(\varepsilon^I)$ for inflection points, in black; $\alpha_G(M) = \alpha_{N,M}(\varepsilon \to \infty)$ for globules, in blue. These values are averaged over N. Shaded regions correspond to $2\sigma_\alpha$ confidence intervals for different sample sizes from 256 (larger intervals) to 2048, calculated by propagating the statistical variance through Eq. (3.14).

independent of M. Thanks to this definition, we obtain a good representation of the phase diagram:

Phase diagram from decimated conformations

For sufficiently large samples, even an extreme decimation with only 3 equally spaced points allows a very accurate reconstruction of the (N, ϵ) phase diagram, as shown in Figure 3.16-b. The critical line $\epsilon^I(N)$ matches that of the complete chain (white line in Figure 3.16-b). Moreover, for $M > 10$, the effects of decimation completely disappear, allowing for a virtually perfect estimation of α.

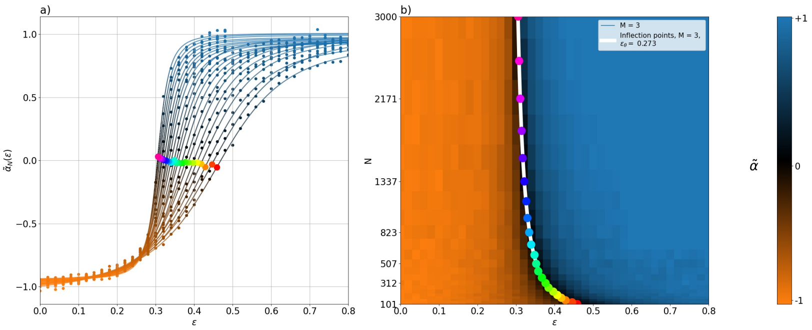


Figure 3.16: **(a)** Renormalized sigmoids $\tilde{\alpha}_{N,3}(\varepsilon)$ (Equation 3.25) for N from 101 to 3000 with a maximal decimation level $M = 3$. **(b)** Phase portrait as reconstructed from a $M = 3$ decimated signal. Red dots are obtained from inflection points as in Figure 3.14. The white line is a fit yielding $\epsilon_\theta = 0.268$. In all graphs, we used a dataset of 16384 configurations for each (N, ε, M) condition.

3.7 CONCLUSION / TAKE-HOME MESSAGES

Let's summarize the key results provided in this chapter. Here, I produced an extensive simulation study of the coil-globule phase transition, concentrating on a particular statistical observable that was never explored in this context before, the equilibrium fluctuations of the Rouse modes, $\langle X_p^2 \rangle$.

In light of the previous chapter, we described and explained the scaling behavior of $\langle X_p^2 \rangle$ across the coil-globule phase transition:

1. In the coil phase, $\langle X_p^2 \rangle$ was found to exhibit a decaying power law scaling with exponent $\alpha \approx -2.17$, reminiscent of the fractal dimension of the (self-avoiding) coil polymer.
2. In the globule phase, $\langle X_p^2 \rangle$ has a segmented shape, reflecting its scale-dependent internal structure. In the low-frequency modes, $p < p^*$, it assumes a constant spectrum $\langle X_p^2 \rangle \sim p^0$, while a power law with exponent $\alpha = -2$ is observed for the high-frequency modes.
3. Finally, in the transition phase, we observed $\langle X_p^2 \rangle$ to continuously evolve from its coil to its globule form.

Crucially, our findings indicate that during the collapse, while the high-frequency modes remain essentially constant, the low-frequency modes undergo significant change, implying that the information about the folding state of the polymer is concentrated in the low-frequency modes.

From here, I introduced an order parameter for the phase transition, named α and defined as the log-log slope in the first two modes of $\langle X_p^2 \rangle$. By simply plotting its value for a large set of (N, ϵ) conditions, I could draw an accurate phase diagram of the finite size coil-globule phase transition, attesting that α is indeed an excellent observable to identify the folding state of a polymer.

Most importantly, unlike previously studied order parameters related to the size scaling of the polymer, α can be computed from a single data set of equilibrium polymer conformations at **a fixed**[31] **length** N.

Finally, given that α is computed from the first two modes, I showed that it can be derived from a very limited amount of information. Namely, using the decimation property of the Rouse modes, I showed that the value of α for a specific polymer of length N, could be computed from the spectrum of the reduced chain constituted of $M < N$ equally spaced points along the original chain. This, however, comes at the cost of some distortion due to aliasing effects, which become more pronounced as the number of points M decreases. Regardless, we found that, even with the information of only 3 points, if aliasing effects were taken into account, the same phase diagram could be reconstructed, attesting to the efficacy of the order parameter even at these extreme decimation conditions. For $M = 10$, I found that aliasing effects essentially disappeared and a perfect estimation of α could be obtained.

In synthesis, by addressing the specific question of assessing the folding state of an interacting polymer, I showcased in this chapter the effectiveness of $\langle X_p^2 \rangle$ in synthesizing the useful information about the polymer in a few low-frequency modes. However, it is important to note that this represents just one example of how this spectral observable can be used. While a homogeneous interacting polymer model can effectively capture chromatin organization at the TAD (Topologically Associating Domain) level, as demonstrated by Lesage and colleagues[32], it's important to recognize that for a comprehensive un-

[31] Remember that the value of R_g for a **single** polymer, without any other data points, **doesn't** allow the assessment of polymer's folding state. The reconstruction of the scaling of R_g is required for that.

[32] A. Lesage, V. Dahirel *et al.* Polymer coil–globule phase transition is a universal folding principle of drosophila epigenetic domains. *Epigenetics & Chromatin*, 12(1), May 2019, [47].

derstanding of higher-order chromatin structures, more sophisticated polymer models are essential. To fully exploit the potential of this method in the study of chromatin organization, it is imperative to characterize a broader range of polymer models relevant to chromatin modeling (heteropolymer models, fractal globule, looped polymers, etc...). For this reason, I believe the **PSD should be considered as an observable to be systematically investigated** when analyzing polymer models, similarly to other metrics such as contact probability or radius of gyration.

In this spirit, in the next chapter, I'll start by using the tools developed so far to study the folding state of chromatin in **experimental data**. Also, I'll apply the spectral method to the different task of detecting if a polymer conformation presents one or several loops, and again apply the method to real chromatin data.

4

Spectral Analysis of Chromatin Conformations

In the previous chapter, I introduced a statistical measure, the Power Spectral Density of a polymer conformation. We applied this measure in the context of analyzing computer-simulated interacting polymer chains. Our analysis demonstrated its utility in identifying the folding state of these chains, due to its effectiveness at isolating their large-scale features, that are most influenced by conformational changes during the transition.

In this chapter, we aim to make use of this observable **on chromatin data**. As established in the previous chapter, computing the PSD of a chromatin domain requires the knowledge of a minimum of 3 equally spaced loci along the domain of interest. In fact, computing $\langle X_p^2 \rangle$ requires an **ordered** sequence of chromatin segment positions. In other words, we need the position of the chromatin segments while also knowing their genomic coordinates, i.e. their position along the DNA chain.

While it is relatively easy, using fluorescent markers, to visualize several chromatin locations simultaneously (Figure 4.1, top), keeping track of the respective genomic location of each segment is extremely challenging and was, until recently, only possible for a very few genomic locations, using multiple label colors.

In 2016, Wang et al.[1], developed the **sequential FISH (seq-FISH) method** allowing the visualization of dozens of genomic locations, while keeping track of the identity of each segment, in thousands of fixed cells at once (Figure 4.1, bottom). In this seminal work, markers were placed every 1 Mbp, probing chromatin structure well above the TAD level. Two years later, the method was refined by B. Bintu et al.[2], leading to a much greater genomic resolution, aiming for an investigation of the internal structure of TADs.

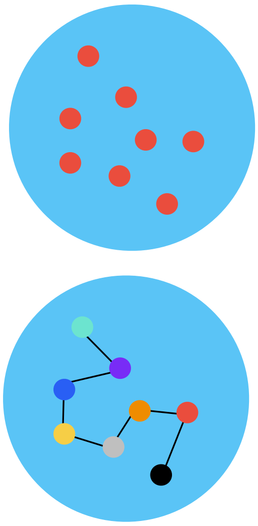


Figure 4.1: Schematic representation of the nucleus during a classic FISH (top) and sequential FISH (bottom) experiments

The structure of this chapter will be as follows. In section 4.1, I'll describe this **experimental seq-FISH method**. In section 4.2 I'll move on to compute and analyze the **power spectral density** of a human chromatin seq-FISH dataset, using the concepts intro-

[1] S. Wang, J.-H. Su *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, 353(6299):598–602, 2016, [82].

[2] B. Bintu, L. J. Mateo *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, 2018, [9].

duced in chapter 2 and chapter 3. As we will see, the computed spectra will prove challenging to interpret from the simple homogeneous model proposed in chapter 3. In this respect, I propose a few **possible interpretations**, and discuss their weaknesses and strengths.

I'll then change my perspective in section 4.3 to section 4.5, where I'll discuss the importance of detecting loops in chromatin data, already mentioned in the introduction, and present another **spectral-based method for detecting loops** in seq-FISH data.

4.1 SEQUENTIAL DNA FLUORESCENT IN-SITU HYBRIDIZATION

4.1.1 *DNA Fluorescent In-situ Hybridization*

DNA fluorescence in-situ hybridization (FISH) is a molecular biology technique used in microscopy experiments to visualize and locate specific DNA sequences within the nucleus. The general idea of this method is to attach to the DNA segment of interest a fluorescent molecule, called a **fluorescent probe**, that can easily be localized by optical microscopy.

The fluorescent probe is a molecular complex made of two pieces: a single-strand DNA segment, called the **probe**, and a fluorescent molecule called the **fluorophore**. Both elements are attached by a covalent bond, through a process called probe labeling. To have the probe specifically attach to the target DNA segment, its sequence is chosen as the **complementary** sequence to the segment of interest.

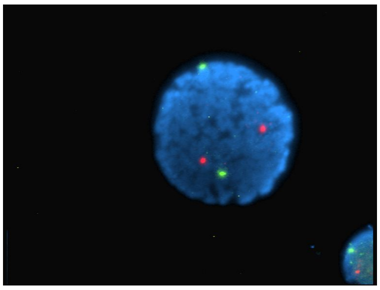


Figure 4.2: human lymphocyte nucleus stained with FISH probes on chromosomes 13 (green) and 21 (red). From Wikimedia Commons, courtesy of user Gregor1976.

Of course, before any intervention, the target DNA strand is already bound to its natural complementary sequence. To introduce the probe, the experimentalist breaks the hydrogen bonds between the base pairs of the target sequence, allowing the probe to take the place of the complementary sequence. This process is called **denaturation** and is performed by subjecting the cell to high temperatures, which has the effect of locally unwinding the double-stranded DNA, creating "denaturation bubbles", i.e. areas of partially separated DNA.

While the DNA is in the denatured state, in a process called hybridization, fluorescent probes are introduced in the nucleus, with the hope that one successfully **hybridizes**, i.e. pair up with the exposed single strand of the target DNA. The denaturation process is then reversed, at least wherever no marked sequences occupy the place of the other strand. In this step, the stability of the DNA double helix is of tremendous importance, as it allows the DNA to recover to a normal form with the markers in place. For added efficiency, probes with both complementary sequences are introduced, such that two probes can attach, one to each strand of the sequence.

After hybridization, all non-hybridized probes are removed in a process called **washing** during which the sample is rinsed under a solution of, typically, ethanol. Note that, the longer the probe, the more probable it is to hybridize and remain attached to the region of interest during the washing process, allowing for a better detection rate.

After the washing process, the sample is ready for detection. The fluorescent probe, which has specifically hybridized to the target DNA sequence, will emit fluorescence when exposed to appropriate wavelengths of light, allowing one to visualize and locate the specific DNA sequences of interest.

4.1.2 *Sequential DNA FISH*

In principle, by targeting different DNA regions, one can visualize numerous genomic locations at once in each round of imaging. However, the challenge lies in distinguishing which illuminated spot corresponds to which genomic location. While employing multi-colored fluorophores, allows for a few loci to be simultaneously imaged, to trace the path of chromatin over large segments, many sequential FISH imaging rounds must be performed in succession, in a process called **sequential FISH (seq-FISH)**.

In this method, in a first process called **fixation**, all elements of the nucleus are glued together by creating covalent cross-links between them. This process is done while the cell is still living, effectively freezing into a life-like conformation.

Then, the region of interest is partitioned into N segments, of genomic length b_{kb}, that determine the **genomic resolution** of the experiment. The N segments are then hybridized all at once by segment-specific **primary probes**. These probes remain attached to the targeted region for the whole time of the experiment. The primary probes aren't labeled with fluorophores but are vessels to which another probe, called a **secondary probe or readout probe**, which is fluorescent, can specifically hybridize.

In this way, each primary probe can be "turned on" individually by introducing its corresponding readout probe. A visualization round is performed to localize the position of the first segment. The secondary probe is then washed out and the process is repeated for each primary probe until the whole region of interest has been imaged. Finally, to obtain the chromatin conformation, the positions of each individual segment are overlaid. Using a microfluidic chamber to administer the different chemicals, this process can be automated and performed on thousands of cells at once.

The final result is a set of **thousands of independent 3-dimensional conformations** of a chromatin region of length Nb_{kb}. The complete conformational information (up to the genomic resolution) offered by seq-FISH makes it possible, for the first time, to compute any ensemble averaged polymer physics observable.

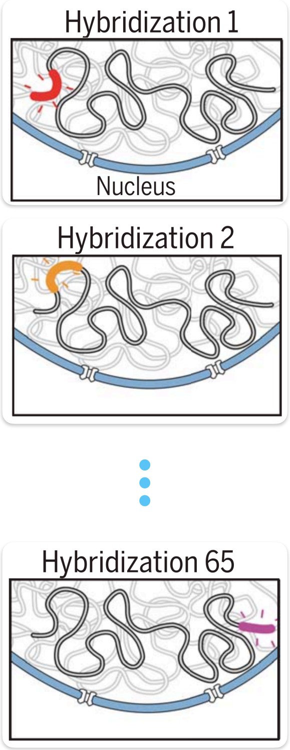


Figure 4.3: Sketch of the sequential FISH hybridization approach. Figure adapted from Ref. [9].

4.2 Assessing the Folding state of Human chromatin seq-FISH data

4.2.1 *Bintu et al. Experimental Data Set*

In this section, we investigate a dataset from the 2018 publication by Bogdan Bintu and co-workers[3]. In this experimental paper, the authors perform seq-FISH on chromosome 21 in different **human** cell lines, during **interphase**.

Two batches of experiments were conducted, aiming for different objectives. In the first, they image the same genomic region in different cell lines, allowing for a comparison of structural properties.

In the second, in the same cell line and region, they compared a wild-type variant with a RAD21-depleted variant. RAD21 is one of the building blocks of **cohesin**, a structural protein complex involved in **loop-extrusion**, which in turn is involved in TAD formation and stabilization. Hence, depletion of RAD21 eventually leads to TADs no longer being visible on the level of the whole population[4]. This is illustrated in the median distance maps for both variants, depicted in Figure 4.4. The cohesin-depleted variant shows none of the TAD and sub-TAD structures visible in the untreated variant.

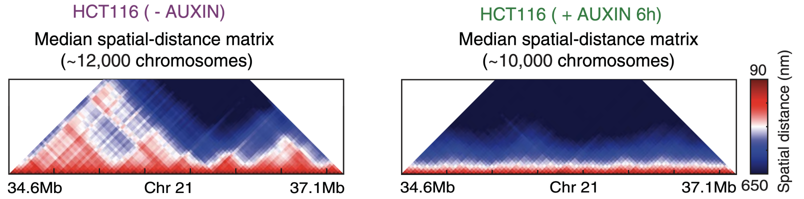


Figure 4.4: Population averaged spatial-distance matrices for in the HCT116 cell line without (left) or with (right) auxin treatment to induce cohesin degradation. Taken from Bintu et al. [9].

In each experiment, **the primary probes were 30kb long** and were placed contiguously along the region of interest, enabling the investigation of sub-TAD architecture while ensuring a high detection rate of around ninety percent[5]. The condition, regions, and cell lines of the datasets are reported in the following table:

Exp. N^{o}	cell line	variant	gen. coord. (Mb)	N^{o} of cells	N^{o} of segments
1	IMR90	WT	28-30	1277	63
2	K562	WT	28-30	13997	63
3	A549	WT	28-30	3941	63
4	HTC116	WT	28-30	1979	63
5	HCT116	WT	34.6-37.1	11631	85
6	HCT116	Auxin	34.6-37.1	9526	85

Table 4.1: List of all the experiments of Ref. [9] with details, and an identification number that will be used in the following.

---

[3] B. Bintu, L. J. Mateo *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, 2018, [9].

[4] S. S. Rao, M. H. Huntley *et al.* A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, 2014, [63].

[5] The method of interpolation used to fill the gaps had no influence on the result.

4.2.2 *Spectral Analysis of the Data*

To compare the experimental data with the simulation results of chapter 3, we need to fix a few physical scales. In chapter 3, I simulated fully-flexible polymers, meaning we assumed the persistence length of the polymer was equal to the bond size. For the data, the persistence length is known to be of the order of 1 kb[6]. This means that **one monomer** in the simulation corresponds to **≈ 1 kb** in the experiments. Therefore, the genomic resolution of 30 kb means that the data is at a higher level of decimation than the simulation. To obtain an **equivalent model** from the simulation, a **decimation of order 30** should be performed, i.e. we should keep one every 30 points in the simulation. The equivalent length of the simulated polymer is given by: $N = \frac{\text{genomic length}}{\text{persistence length}}$ – which is $N = 2000$ or $N = 2500$ depending on the dataset.

Also, we should acknowledge that the data is inevitably **noisy**, due to the various sources of imprecision of the measurement. This noise can reasonably be considered as Gaussian white noise, meaning it is expected to **add a positive constant to the spectrum**.

Let's then start by calculating the **PSD of the experimental datasets** listed in Table 4.1. The resulting spectra are all plotted in Figure 4.5. The general trend for all experiments is a power law scaling with exponents ranging from ≈ -1.7 and ≈ -1.4 (with one outlier at -1.1). In the high-frequency modes, the spectrum deviates from the power law behavior as expected due to aliasing effects.

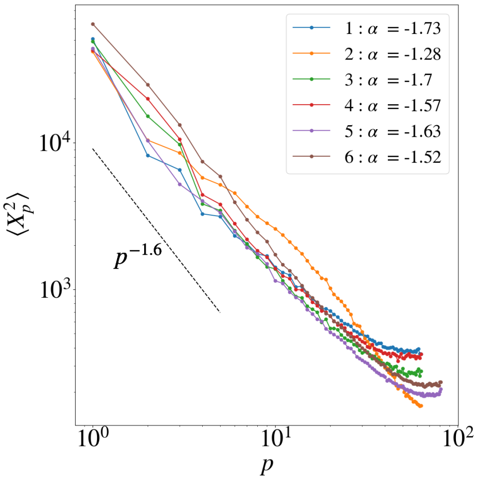


Figure 4.5: DSP for all the datasets of Ref. [9] listed in Table 4.1.

The first observation we can make is that the spectra appear different from the coil-globule spectra studied in the previous chapter: Indeed, in the previous chapter, the only power law spectra we observed were for the coil phase, with exponent -2.17 or high-frequency regime of the globule, with exponent -2 (ideal chain behavior). Deviations from this behavior were expected due to decimation, but their effect was not strong enough to justify the observed exponents: Here, we observe **a power low behavior with a rather lower exponent**.

Due to the polymeric nature of chromatin, we indeed expect that at some small enough length scale, the chain exhibits either a Gaussian or self-avoiding behavior. Hence, our first observation is that this behavior is not visible, implying that much of the fine structure of the chromatin fiber is completely overlooked at this genomic scale.

This raises the question of how to interpret the obtained spectra. I propose **hereafter some possible explanations** for the observed exponents.

4.2.3 *Crumpled Globule model*

Another model, conceptually different from the coil-globule model, is often proposed to describe chromatin architecture: the **crumpled globule - or fractal globule - model**. This model (already mentioned in the introduction) proposes that chromatin organizes as an out of equilibrium meta-stable state. In this state, the predicted confor-

[6] A. Lesage, V. Dahirel *et al.* Polymer coil–globule phase transition is a universal folding principle of drosophila epigenetic domains. *Epigenetics & Chromatin*, 12(1), May 2019, [47].

mation is similar to the equilibrium globule, in that it is a space-filling curve in a roughly spherical volume, except that, unlike its equilibrium counterpart, it is *completely free of knots*. Its unusual topology drastically changes its inner structure, making its conformations statistically fractal. The fractal exponent predicted for the fractal globule is $d_F = 3$, like for a solid 3D object, because of its space-filling character[7].

Following the same reasoning as for the coil polymers (subsection 3.3.2), we show that the fractal nature of its conformations implies the following scaling law for the **PSD of the fractal globule**,

$$\langle X_p^2 \rangle \sim p^{-(1+2\nu)}, \tag{4.1}$$

where $\nu = 1/d_F$: Transposing the argument directly to the fractal globule model provides a PSD power law exponent of $-(1+\frac{2}{3}) \approx -1.66$. This suggests that a crumpled globule organization for chromatin may indeed offer a plausible explanation for the observed power-law in the PSD.

Nevertheless, the very existence of the fractal globule has been questioned, due to its instability[8], and especially at smaller length scales, $<$ 1Mb, where it is unequivocally expected to exhibit Gaussian behavior[9]. In particular, Schram and co-workers show by convincing Monte Carlo simulations of collapsing polymers that

> the very property of being disentangled also brings about the immediate destruction of the fractal state,

inducing a rapid equilibration toward a state that shares the main scaling features of the equilibrium globule[10].

Hence, while the fractal globule is indeed still a (highly debated) contender to describe the higher order structure of chromatin $>$ 1 Mb, it is, in my opinion, **unlikely** that it is a pertinent model for describing the structure **at the scale probed in this experiment**, i.e. sub-TAD to a few TADs, due to its high instability on small length scales. Moreover, as we'll see in the second part of this thesis, the fractal globule model predicts the wrong dynamics for the motion of chromatin[11,12].

4.2.4 *Decimated, critical polymer for TAD-depleted chromatin*

If the fractal polymer model gives a possible interpretation for the observed exponent, this exponent can also be explained in the context of the interacting polymer model presented in the previous chapter. Indeed, remember that, in the transition phase, the log-log slope in the first modes lies between -2.17 and 0. Then, if only the first modes were visible, essentially any intermediate slope would be observed. Hence, the exponent of -1.7 can be explained with **a transition polymer model at a high level of decimation**.

To check the pertinence of this effect, a polymer of length $N = 2194$ and $\epsilon = 0.3$, i.e. at the transition (see Figure 3.14), decimated at

---

[7] L. A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research*, 19:37–51, 2011, [55].

[8] R. D. Schram, G. T. Barkema *et al.* On the stability of fractal globules. *The Journal of Chemical Physics*, 138(22):224901, 06 2013, [67].

[9] A. Chertovich and P. Kos. Crumpled globule formation during collapse of a long flexible and semiflexible polymer in poor solvent. *The Journal of chemical physics*, 141:134903, 10 2014, [13].

[10] R. D. Schram, G. T. Barkema *et al.* On the stability of fractal globules. *The Journal of Chemical Physics*, 138(22):224901, 06 2013, [67].

[11] M. Tamm, L. Nazarov *et al.* Anomalous diffusion in fractal globules. *Physical review letters*, 114, 04 2014, [75].

[12] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023, [11].

order 30 is presented in Figure 4.6 (faded dashed blue line) along with experiment 6.

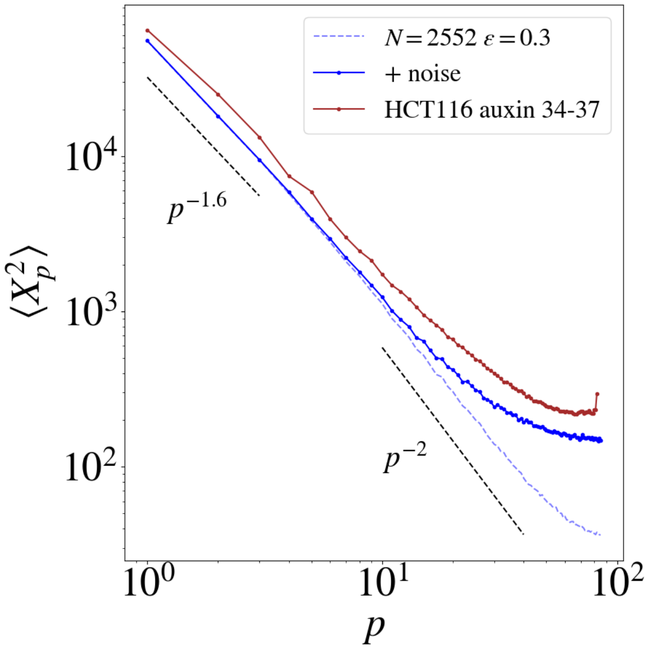


Figure 4.6: PSD for a simulated transition polymer of length $N = 2552$ and $\epsilon = 0.3$ decimated to order 30 (blue) versus the PSD obtained from the RAD21 depleted HCT116 cells in region $34.6Mb - 37.1Mb$ (brown).

The good agreement between the experimental curve and the results of our decimated, noisy transition polymer allows us to draw some first conclusions: We see that the PSD of the simulated transition polymer indeed appears to exhibit the **correct power law behavior** in the large frequency modes. In the high-frequency modes, the decimated simulation is seen to adopt a higher exponent, close to -2, hence drops quicker than the experimental data. However, if we **add Gaussian white noise** to the simulation (blue curve) the added constant raises the high frequencies, matching nicely the behavior of the **auxin-treated (RAD21-depleted) experimental curve**, i.e. experiment 6.

Note that, to obtain Figure 4.6, the variance of the white noise was adjusted "by hand" to fit the high-frequency modes of the experimental PSD. The value that fits best is essentially identical to the variance of the bond length in the decimated simulation. Further study would be interesting to verify the **precise origin of this noise** and how it relates to the details of the experiment. In particular, we could try to link the noise to the fact that in FISH the entire subchain corresponding to 30 monomers is imaged.

4.2.5 *Single cell TAD-like Structures After Cohesin Depletion*

Since depleting RAD21 (equivalently cohesin) results in TADs disappearing at the population level (see Figure 4.4), it can be expected that TAD structures should be completely absent at the single cell level as well. However, one of the major findings of Bintu et al. is precisely that **TAD-like structures remain visible in *single cells* after cohesin depletion**. To explain the disappearance of TADs at the population level the authors specify that:

> What was notably different in the absence of a functional cohesin complex was that the positions of these domain boundaries became largely uniformly distributed along the genomic coordinate and **no longer exhibited preferential positioning at CTCF and cohesin sites**[13] as observed in the presence of cohesin. These results indicate that cohesin is not required for the maintenance of TAD-like structures in single cells and that the role of cohesin in the formation of ensemble TADs is to establish preferred genomic boundaries for the single-cell domains.

To illustrate this, in Figure 4.8 a) and b), I reproduce the 4 first experimental single-cell conformations, along with their respective distance maps from the cohesin depleted. The presence of diagonal blocks indeed confirms that contiguous genomic segments display preferential spatial proximity at the single-cell level.

Interestingly, these structures can again be explained by the **critical nature** of the chromatin fiber. Indeed, as mentioned in section 3.5, in the critical phase self-attracting polymers display **strong conformational fluctuations**, as indicated by the diverging relative fluctuations in the first spectral modes (see figure Figure 3.13). These fluctuations are due to the polymer not having quite enough attractive energy to fully collapse, leaving it in a **hybrid state in between the coil and the globule states**. At the single conformation level, this shows as the polymer being **partially collapsed**, i.e. presenting both collapsed globular and coil-like subchains. To give a more intuitive visualization of such conformations, an animation showing the dynamics of a critical polymer is shown in Figure 4.7.

Figure 4.7: Animation of a critical polymer simulated using Langevin dynamics. A thorough analysis of these simulations will be done in the second part of the thesis. Requires *Adobe Acrobat Reader* to be read. If you don't have this software, you can also click this *link*

These claims are justified in Figure 4.8, where conformations from a (simulated) critical polymer of length $N = 2154$ and $\epsilon = 0.3$ are shown (c), along with their order 30 decimation (d) and the corresponding distance map (e). Clearly, the **simulated single-cell contact maps** present similar features to the experimental ones, with the **appearing of transient TAD-like structures**. The probability for each genomic position to be a single-cell domain boundary is of course, for the simulations, uniform, in agreement with Ref.[14] for Auxin-treated cells.

[13] Emphasis added.

[14] B. Bintu, L. J. Mateo *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, 2018, [9].

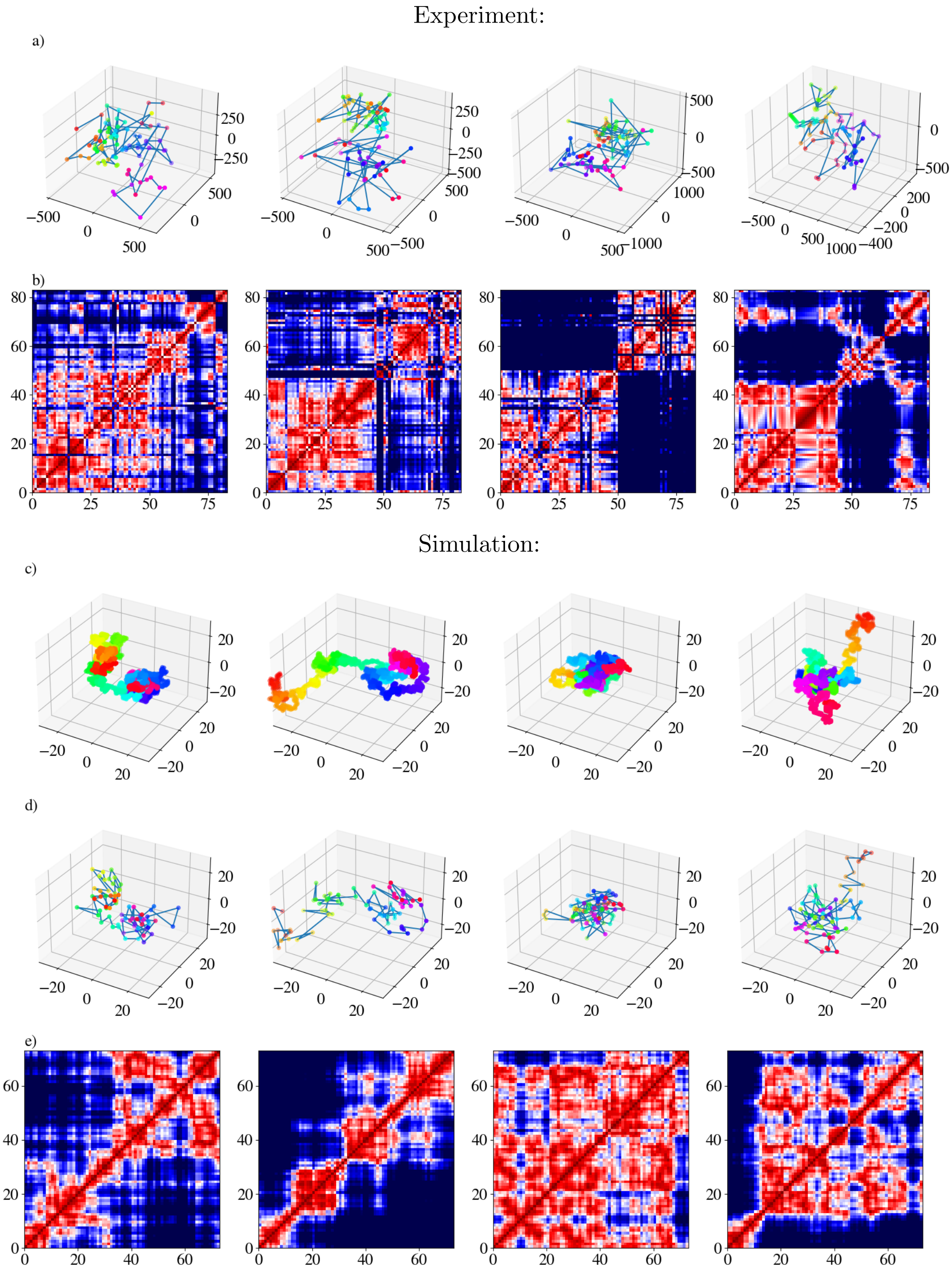


Figure 4.8: (a) 4 single-cell chromatin conformations with auxin treatment taken from the experimental dataset. (b) the corresponding single-cell distance matrices. (c) 4 conformations of a computer-simulated critical polymer of length $N = 2154$ and $\epsilon = 0.3$. (d) the corresponding order 30 decimation conformations with the added noise. (e) the corresponding distance matrices.

4.2.6 *Criticality explains TAD stabilization*

As briefly mentioned in the introduction, CTCF-mediated loops are known to be responsible for stabilizing TADs. A potential explanation for this can be put forward from the partially collapsed, critical picture of chromatin. Indeed, being in a critical state, the slightest perturbation may induce a local collapse of the chain into a globular state. Consequently, if a loop is formed within the chromatin chain, the subchain forming the loop would collapse, while the surrounding chromatin would remain in the critical state. This would result in a higher proximity among the elements enclosed within the loop than outside the loop.

To test this idea, we studied the collapsing thermodynamics of simulated on-lattice **interacting circular** polymer. This work was done by Michaël Liefsons, during his master's internship in our lab. This work was a first step towards the work I will present in the next section. I won't present the whole analysis here, but simply the final result. Using a similar spectral method, adapted to circular conformations, we could define an order parameter, similarly to the one defined in chapter 3. From this order parameter, we could draw the phase diagram of the coil-globule phase transition of a circular interacting polymer, which is shown in figure Figure 4.9. The critical line for the circular and linear polymer is drawn in white and yellow respectively.

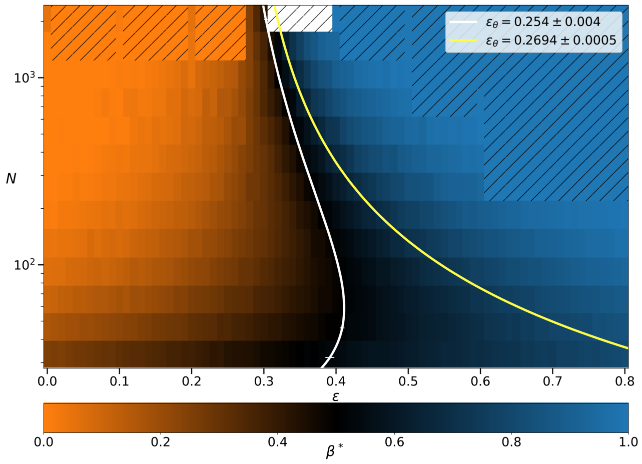


Figure 4.9: Phase portrait of the coil–globule transition for looped polymers. The yellow line is the critical curve for the circular conformation and the white line for the linear conformations.

Crucially, we notice that the looped critical curve is located at a lower energy than the linear critical curve, meaning **collapsing a looped chain requires less energy** than a linear chain. In fact, the critical line for the linear chain is in the globule phase of the looped phase diagram. This implies that **a critical linear chain, when looped, can transition to a collapsed state**.

4.2.7 *Heterogeneity induced by the presence of TADs*

The other experimental curves are less smooth than the auxin-treated experiment. Yet, **tuning** by hand **the value of** ϵ **and the noise** allows for the general trend of all experimental curves to be more or less well reproduced.

Nevertheless, we are aware that, in these experiments, due to the presence of cohesin, TAD structures are stabilized and are present at the population level (see the average distance map in Figure 4.4).

Consequently, we don't expect a **homogeneous** self-attracting polymer model to perfectly reproduce the chromatin structure in the non-treated data sets. The effect of a **heterogeneous folding** of the chain has not been explored in this thesis. However, it could indeed be a valuable matter **to address in future work**, and might play a crucial role in understanding the more complex spectra observed in the WT variants.

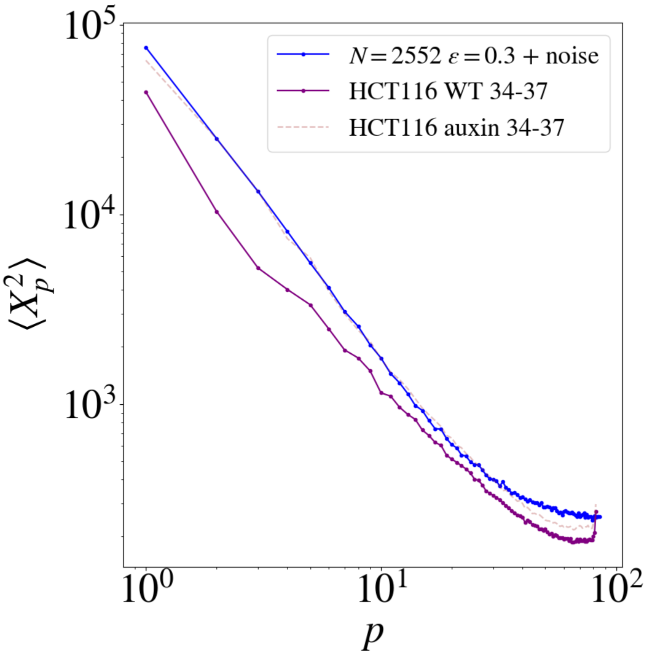


Figure 4.10: PSD for a simulated transition polymer of length $N = 2552$ and $\epsilon = 0.3$ decimated to order 30 (blue) versus the PSD obtained from the WT HCT116 cells in region $34.6Mb - 37.1Mb$ (purple). The RAD21 depleted PSD

More generally, it's clear that the precise interpretation of these spectrums warrants further investigation. Alternate regions of different degrees of compaction are probably one of the causes. However, more sequence-specific effects are probably involved, in particular in regions containing regulatory elements. Furthermore, and this will serve as a transition to the next section, one very important aspect of chromatin organization has been completely overlooked in this discussion and plays an important role in the observed WT spectra: chromatin's ability to form loops through loop extrusion, which is precisely what auxin inhibits. Indeed, we can expect the average spectrum of **looped conformations** to be greatly affected compared to their linear equivalent, potentially explaining some of the roughness observed in WT spectra. This idea is investigated in the following section.

For the moment, let me resume the main results of thissection 4.2:

*The critical nature of the polymer explains features present in **cohesin depleted** data*

The take-home message of this section is therefore that human chromatin conformations of Ref. [9], on genomic scales ≈ 30 kb to ≈ 2 Mb and when depleted of TAD structures, is well described by a **homogeneous critical interacting polymer**, in line with the finding of Lesage et al. [47].

Indeed,

- The PSD of the simulated transition polymer, decimated according to the experiment resolution and with the addition of white noise, appears to reproduce the behavior of the **auxin-treated (RAD21-depleted) experimental curve**.
- Due to the large fluctuations characterizing the transition polymer, the **simulated single-cell contact maps** present similar features to the experimental ones, with the **appearing of transient TAD-like structures**.
- Furthermore, a comparison of the behavior of circular and linear polymers suggests that, if a loop is formed within the chromatin chain, the subchain forming the loop would collapse, hence stabilizing the corresponding TAD.

For the **wild type**, a decent fit to data is obtained as well from the same model in similar conditions, but the curves present finer details that aren't accounted for by this homogeneous model, letting one think that, with a critical interacting polymer model as basis, one should add **additional features** to make it more realistic. The key elements here would include

(i) incorporating specific interactions among monomers within TADs,

(ii) accounting for variations in local compaction levels within the chromatin structure,

(iii) studying the effect of loop formations on the spectral signature.

4.3 LOOPS AND LOOP EXTRUSION

Remark: *The work presented in the next sections of this chapter was a collaborative effort with a very talented master's student called Michaël Liefsoens. Michael was recruited in Maria's team during my second year as a Ph.D. student to extend the spectral method I developed to characterize different polymer topologies, with a specific focus on looped polymers. In this research, my role was more of a mentor than the primary investigator.*

The following sections are adapted from the paper

Spectral-based detection of chromatin loops in multiplexed super-resolution FISH data
by Michaël Liefsoens, Timothy Földes and Maria Barbi

submitted to *Nature Methods* and available in arXiv.

Loop formation is central to understanding chromatin architecture and its functional role. During mitosis, chromatin adopts a compact structure composed of loops, forming a rod-like configuration [15]. SMC (structural maintenance of chromosome) proteins like **condensins** and **cohesins** play a pivotal role in organizing these loops [16]. Recent research reveals that loop formation, mediated by proteins such as CCCTC-binding factor (**CTCF**) and cohesin, is also critical in interphase, for gene regulation by facilitating interactions between distant enhancers and promoters in mammals [17,18], *Drosophila* [19], and yeast [20]. Furthermore, cohesin-dependent loops are involved in the segmentation of interphase chromosomes into TADs which are often delimited by CTCF binding sites. Depletion of CTCF disrupts both TAD loops and insulation of neighboring TADs [21]. Consequently, the identification of chromatin loops has become central to unraveling the complexities of gene regulation and understanding the spatial organization of the genome. Key questions arise regarding loop formation mechanisms, their prevalence, determinants of their position and sizes, and biological functions.

The **loop extrusion** mechanism[22], primarily involving SMC family proteins like cohesin (in interphase) and condensin (in metaphase), can explain loop formation. Cohesin and CTCF enable loop extrusion by binding to DNA, after which they act as motors, sliding in opposite directions and enlarging the loops by pulling along the chromatin fibers[23]. Looping by SMC complexes is observed in various cell types, including mammalian and bacterial cells[24]. As insulator proteins, CTCF and cohesin regulate chromatin loop stability, probably as a 'dynamic complex' that frequently breaks and reforms throughout the cell cycle[25].

In high-throughput genomic techniques like Hi-C [26], stable loops manifest as **isolated, off-diagonal points** in contact maps. Data

[15] J. R. Paulson, D. F. Hudson *et al.* Mitotic chromosomes. *Seminars in Cell & Developmental Biology*, 117:7–29, 2021, [59].

[16] J. R. Swedlow and T. Hirano. The making of the mitotic chromosome: Modern insights into classical questions. *Molecular Cell*, 11(3):557–569, 2003, [72].

[17] S. S. Rao, M. H. Huntley *et al.* A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, 2014, [63].

[18] M. A. Karpinska and A. M. Oudelaar. The role of loop extrusion in enhancer-mediated gene activation. *Current Opinion in Genetics & Development*, 79:102022, 2023, [39].

[19] S. M. Espinola, M. Götz *et al.* Cis-regulatory chromatin loops arise before tads and gene activation, and are independent of cell fate during early drosophila development. *Nature Genetics*, 53(4):477–486, 2021, [21].

[20] L. Costantino, T.-H. S. Hsieh *et al.* Cohesin residency determines chromatin loop patterns. *eLife*, 9:e59889, nov 2020, [15].

[21] E. P. Nora, A. Goloborodko *et al.* Targeted degradation of ctcf decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, 169(5):930–944.e22, 2017, [56].

[22] E. Alipour and J. F. Marko. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic acids research*, 40(22):11202–11212, 2012, [4].

[23] G. Fudenberg, M. Imakaev *et al.* Formation of chromosomal domains by loop extrusion. *Cell Reports*, 15(9):2038–2049, 2016, [26].

[24] E. J. Banigan, A. A. van den Berg *et al.* Chromosome organization by one-sided and two-sided loop extrusion. *eLife*, 9:e53558, apr 2020, [6].

[25] A. S. Hansen, I. Pustova *et al.* Ctcf and cohesin regulate chromatin loop stability with distinct dynamics. *elife*, 6:e25776, 2017, [34].

[26] E. Lieberman-Aiden, N. L. van Berkum *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009, [49].

analysis tools detecting DNA loops in contact maps, based on contact count enrichment or specific patterns, are available [53] [64] [65].

However, Hi-C methods lack the ability to reconstruct the polymer's spatial trajectory, only quantifying contact frequencies between monomers. This limitation is overcome by the seq-fish data presented in the subsection 4.1.2. However, the most frequent approaches to analyze this new data are based on the reconstruction of *distance* maps, then interpreted as contact maps [27]. However, this approach restricts the information to a level already obtainable with previous techniques and, most importantly, doesn't allow for the detection of loops.

However, this approach restricts the information to a level already obtainable with previous techniques.

Innovative methods are clearly needed to fully exploit this new data. With Michaël Liefsoens, we looked for the possibility of **characterizing chromatin loops through a spectral representation** of chain configurations, thereby exploiting the whole information of chain 3D spatial arrangement offered by seq-FISH methods.

4.4 THE Λ-PLOT FOR LOOP DETECTION

4.4.1 *Comparing PSD for looped and non-looped fBm-based conformations*

As a first step, we wanted to extend the PSD analysis to circular polymers, to examine the impact of looping on the spectrum. We employ a minimal, yet instructive, model of polymer configurations represented as 3D correlated random walks $\boldsymbol{\gamma}_n$, using fractional Brownian motion (fBm) of Hurst coefficient H, already introduced in section 2.3.

Following Ref. [28] and as detailed in the following box, we define a looped fBm as

$$\boldsymbol{\lambda}_n = \boldsymbol{\gamma}_n - \mathcal{B}_n^{(H)}\,\vec{R}\,: \tag{4.2}$$

here, $\vec{R} = \boldsymbol{\gamma}_N - \boldsymbol{\gamma}_1$ represents the fBm end-to-end vector and

$$\mathcal{B}_n^{(H)} = N^{-2H}\,C_{\gamma\gamma}(n, N) \tag{4.3}$$

is the appropriate **bridge function** needed to connect the two ends of the fBm to construct an fBm loop.

For our simplified fBm model, the PSD of the looped chain can be obtained **analytically**. Thanks to the linearity of the DCT, the difference between looped and linear fBm is indeed simply the DCT of the bridge function $\mathcal{B}_H\,\vec{R}$. The symmetry properties of this function then ensure that

(i) **the even modes for looped fBm remain asymptotically unchanged compared to those of the corresponding non-looped**; and

(ii) the odd modes systematically decrease, with the extent of reduction diminishing as the mode number p increases[29].

These results are proven in the following box.

---

[27] L. Lee, H. Yu *et al.* SnapFISH : a computational pipeline to identify chromatin loops from multiplexed DNA FISH data. *Nature Communications*, 14(1):4873, 2023, [46].

[28] D. Gasbarra, T. Sottinen *et al.* Gaussian bridges. In F. E. Benth, G. Di Nunno *et al.*, editors, *Stochastic Analysis and Applications*, pages 361–382, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg, [28].

[29] This property is a general consequence of the condition that the first and last monomer coincide, and thus applies to any looped conformation.

Spectrum of a looped fBm:

We start from the definition of a looped random walk $\vec{\lambda}$, equation (4.2), and calculate the corresponding PSD, here noted $\left\langle X_p^2(\vec{\lambda})\right\rangle$. Analogously, we will denote the PSD of the unlooped $\vec{\gamma}$ by $\langle X_p^2(\vec{\gamma})\rangle$. The linearity of the DCT gives that the PSD of $\vec{\lambda}$ is given as

$$\frac{\left\langle X_p^2(\vec{\lambda})\right\rangle}{\langle X_p^2(\vec{\gamma})\rangle} = 1 - \frac{1}{\langle X_p^2(\vec{\gamma})\rangle}\left(\mathrm{DCT}_{(\mathcal{B}_h(n;0,N))_{n=1,\ldots,N}}(p)\right)^2 \langle \vec{R}^2\rangle : \tag{4.4}$$

it is therefore the DCT of the bridge function $(\mathcal{B}_h(n;0,N))_{n=1,\ldots,N}$ that determines the difference in PSD between the looped and non-looped fBm. By definition, we have for $p > 0$

$$\begin{aligned}
\mathrm{DCT}_{(\mathcal{B}_h(n;0,N))_{n=1,\ldots,N}}(p) =& \frac{1}{2N^{2H+1}} \sum_{n=1}^{N} \left(n^{2H} - (N-n)^{2H}\right) \cos\left(\frac{p\pi}{2N}(2n-1)\right)\\
=& \frac{1}{2}\sum_{n=1}^{N} \frac{1}{N}\left(\left(\frac{n}{N}\right)^{2H} - \left(1-\frac{n}{N}\right)^{2H}\right) \cos\left(p\pi\left(\frac{n}{N}-\frac{1}{2N}\right)\right)\\
\approx& \frac{1}{2}\int_0^1 \left(x^{2H} - (1-x)^{2H}\right) \cos(p\pi x)\, dx.
\end{aligned} \tag{4.5}$$

where we have converted the summation into an integral, valid for $N \gg 1$. Note that $1/N$ of the summation became the volume element in this integral.

Denoting the integrand of equation (4.5) by the function $g_p(x) = \left(x^{2H} - (1-x)^{2H}\right)\cos(p\pi x)$, we note its symmetry: $g_p(1-x) = (-1)^{p+1} g_p(x)$. Consequently, the integral becomes

$$\int_0^1 g_p(x)\, dx = \int_0^1 g_p(1-y)\, dy = (-1)^{p+1}\int_0^1 g_p(y)\, dy,$$

after a change of variables $y = 1 - x$. From this, we can deduce $(1 + (-1)^p)\int_0^1 g_p(x)\, dx = 0$, or

$$\int_0^1 g_p(x)\, dx = 0 \quad \text{for } p \text{ even :}$$

hence, we can conclude that $\left\langle X_p^2(\vec{\lambda})\right\rangle = \langle X_p^2(\vec{\gamma})\rangle$ for p even, i.e. the even modes of looped and non-looped (infinite length) fBm are the same. This is a direct consequence of the symmetry of the bridge function.

Let us now consider a general ideal circular signal $\vec{x}$. Then, the first point x_0 of the signal $\vec{x}$ is equal to the last point x_{N-1}. By the symmetry of the DCT operation, it follows that

$$\sum_{\substack{p=1\\ p\text{ odd}}}^{N-1} X_p \cos\left(\frac{p\pi}{2N}\right) = 0. \tag{4.6}$$

This constraint states that the weighted sum of the odd modes should be zero, and is of topological nature. So, if the first mode is large, the other odd modes have to compensate for this by being small. On average, this leads to the lowering of all the odd modes. Since $\cos\left(\frac{p\pi}{2N}\right)$ is a decreasing function in p, the first mode has the most effect on satisfying this constraint, and goes down the most, relatively speaking.

The behavior of the PSD for non-looped and looped fBm polymer configurations, is visually depicted in Figure 4.11. Interestingly, the difference between looped and non-looped configurations primarily impacts the first modes, emphasizing the crucial role of large-scale features in defining the polymer structure, as already discussed.

4.4.2 *Log-spectral ratio* $\Lambda(\vec{x})$

These spectral features offer a method to distinguish between looped and non-looped configurations. Consider indeed a statistical ensemble of 3D signal realizations $\vec{x}_n$. We introduce the log-spectral ratio $\Lambda(\vec{x}_n)$ for $\vec{x}_n$, defined as the *logarithmic* difference between the observed amplitude of the first mode and the amplitude and the amplitude predicted on the basis of a power-law extrapolation from the second and fourth modes. Some manipulation (detailed in the next box) yields

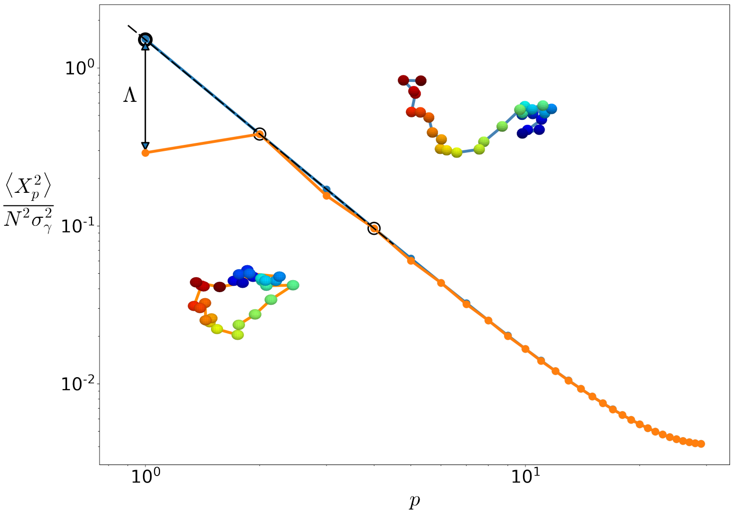


Figure 4.11: Theoretical PSD for an $H = 0.5$ fBm $\vec{\gamma}_n$ (blue) and the corresponding looped $\vec{\lambda}_n$ (orange). Snapshots show one specific conformation before (upper) and after (lower) looping. $\Lambda(\vec{x})$ is the difference between the observed first mode for a looped conformations and the *expected* first mode extrapolated from the second and fourth modes (see black dotted line and circles).

the following **expression for the log-spectral ratio**:

$$\boxed{\Lambda(\vec{x}_n) = \log\left(\frac{\langle \vec{x}_2^2 \rangle^2}{\langle \vec{x}_1^2 \rangle \langle \vec{x}_4^2 \rangle}\right)}. \tag{4.7}$$

Figure 4.11 provides an illustration of this definition. Based on our fBm model, we can demonstrate that the log-spectral ratio for a **non-looped** random walk **scales as** N^{-2} for $N \to \infty$. In contrast, for a **looped fBm**, it converges to a **finite limit** of approximately 1.66, which clearly distinguishes the two configurations[30].

Definition of $\Lambda(\vec{x})$

Since the even modes $\langle X_{2p}^2 \rangle$ are expected to be the same for a looped and a non-looped polymer, we access the expected (linear) coil power law of the low modes by fitting the second and fourth mode. By extrapolating to $p = 1$, we then find the expected outcome for a non-looped polymer, and compare it to the actually observed first mode. Calculating this explicitly, we find

$$\begin{aligned}
&\left[\frac{\log\left(\langle X_4^2 \rangle\right) - \log\left(\langle X_2^2 \rangle\right)}{\log(4) - \log(2)} \left(\log(1) - \log(2)\right) + \log\left(\langle X_2^2 \rangle\right)\right] - \log\left(\langle X_1^2 \rangle\right) \\
&\qquad = -\log\left(\langle X_4^2 \rangle\right) + 2\log\left(\langle X_2^2 \rangle\right) - \log\left(\langle X_1^2 \rangle\right) = \log\left(\frac{\langle X_2^2 \rangle^2}{\langle X_1^2 \rangle \langle X_4^2 \rangle}\right),
\end{aligned}$$

which is exactly how we defined the log-spectral ratio $\Lambda(x)$ in equation (4.7).

For a random walk $\vec{u}$ of variance σ^2 and length N, we can plug in the spectrum $\langle U_p^2 \rangle = \frac{\sigma^2}{8} \frac{1}{N \sin^2\left(\frac{p\pi}{2N}\right)}$ into the definition of Λ to find

$$\Lambda(\vec{u}) = \log\left(\frac{\cos^2\left(\frac{\pi}{N}\right)}{\cos^2\left(\frac{\pi}{2N}\right)}\right) = -\frac{3\pi^2}{4}\frac{1}{N^2} + \mathcal{O}\left(\frac{1}{N^4}\right).$$

For a looped random walk $\vec{l}$, the spectrum

$$\frac{\langle L_p^2 \rangle}{\langle U_p^2 \rangle} = \begin{cases} 1 & \text{if } p \text{ even} \\ 1 - 2\left(N \tan\left(\frac{p\pi}{2N}\right)\right)^{-2} & \text{if } p \text{ odd} \end{cases}$$

gives rise to the following log-spectral ratio:

$$\Lambda(\vec{\ell}) = \Lambda(\vec{u}) - \log\left(\frac{\langle L_1^2 \rangle}{\langle U_1^2 \rangle}\right) = \log\left(\frac{\pi^2}{\pi^2 - 8}\right) + \mathcal{O}\left(\frac{1}{N^2}\right) \approx 1.66 + \mathcal{O}\left(\frac{1}{N^2}\right).$$

---

[30] For finite chains, we compute the absolute difference between the spectra of looped and non-looped random walks, and then normalizing it by the same difference at infinity (= 1.66). This results in a discriminability level ranging between 0 and 1 that can be calculated analytically and converges extremely fast: having $N > 6$ is sufficient to achieve a 90 percent discriminability level; $N > 20$ guarantees a 99 percent discriminability level.

To ensure the robustness and applicability of the $\Lambda(\vec{x}_n)$ definition for signals with varying degrees of correlation, we calculate and display in Figure 4.11 the PSD of **fBm signals with different Hurst exponents** H. Clearly, the behavior theoretically described above and shown in Figure 4.12 is always observed, regardless of the value of H.

The log-spectral ratio $\Lambda(\vec{x}_n)$ proves therefore to be a robust observable that allows us to determine whether a polymer is in a linear or looped configuration. However, to investigate the presence of loops in chromosomes implies two additional issues:

First, chromatin domains are expected to be near the coil-globule transition[31] and exhibit more or less collapsed, globule-like conformations, depending on epigenetics and transcription activity[32,33]. Therefore, it is crucial to verify whether the log-spectral ratio **remains reliable across the coil-globule transition**. We then applied the log-spectral ratio to the Monte Carlo simulations of a cubic lattice self-avoiding walk introduced in chapter 3. Linear and circular polymers were simulated separately, with reptation moves in the former case and Crankshaft rotation, wedge flip, and kink-translocation techniques [34] in the latter, which enhanced simulation efficiency. For the circular polymer, the initial configuration was obtained by the *growing SAW's* algorithm outlined in the same Ref. [80]. Spectra were then estimated and compared for linear and looped polymers across a range of ε values from 0 to 0.5. As shown in Figure 4.13, the difference between the looped and non-looped configurations of the simulated polymers reproduces the expected behavior.

Second, chromatin **loops can vary in size and position** along the chromosome. Consequently, we need to adapt our approach to this more general case.

4.4.3 *The Λ-Plot*

We can introduce an **internal loop** within a random walk by extending the procedure outlined in equation (4.2) to an inner segment. This enables us to generate sets of fBm-based polymer configurations $\{\vec{x}_n\}$ that incorporate one or more internal loops. These loops are defined by their positions ι and lengths η, meaning that monomers $\iota - \eta/2$ and $\iota + \eta/2$ are brought together.

We used these synthetic configurations with internal loops to develop and validate a novel loop-detection technique, known as the **Λ-Plot** and based on the computation of the log-spectral ratio. For each set of N-length signals $\{\vec{x}_n\}$, we consider **all the sub-signals** of length η, defined as $\{\vec{x}^{(\iota,\eta)}\} = (x_{\iota-\eta/2} \dots x_{\iota+\eta/2-1})$. We calculate the log-spectral ratio $\Lambda(\vec{x}^{(\iota,\eta)})$ for each of these sub-signals and represent the results on a color-scale on the plane (ι, η). Figure 4.14 shows typical examples of the expected outcomes when identifying a single inner loop, and compare these results with corresponding distance maps and relevant spectra of sub-polymers.

As shown in Figure 4.14, Λ-plots show a distinct maximum indicating the presence of a loop. A careful inspection reveals that the ι

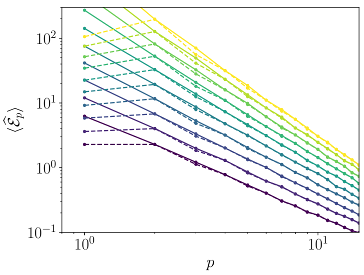


Figure 4.12: Estimated PSD for looped and non-looped fBm signals with varying Hurst exponents ($H = 0.3, 0.35, \dots 0.75$), each from samples of 2000 signals of length $N = 512$.

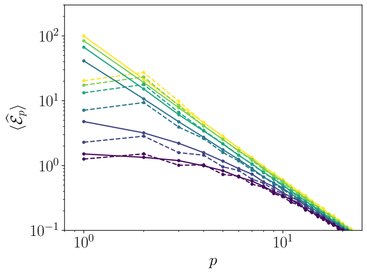


Figure 4.13: Estimated PSD of self-interacting looped and non-looped polymers for $\varepsilon = 0, 0.1, 0.2, 0.3, 0.4, 0.49$ (from samples of 20000 conformations for a $N = 512$ polymer simulated by th on-lattice Monte Carlo of chapter 3.

---

[31] A. Lesage, V. Dahirel *et al.* Polymer coil–globule phase transition is a universal folding principle of drosophila epigenetic domains. *Epigenetics & Chromatin*, 12(1), May 2019, [47].

[32] A. N. Boettiger, B. Bintu *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586):418–422, 01 2016, [10].

[33] Q. Szabo, D. Jost *et al.* Tads are 3d structural units of higher-order chromosome organization in *Drosophila*. *Science Advances*, 4(2):eaar8082, 2018, [74].

[34] T. Vettorel, S. Y. Reigh *et al.* Monte-carlo method for simulations of ring polymers in the melt. *Macromolecular Rapid Communications*, 30(4-5):345–351, 2009, [80].

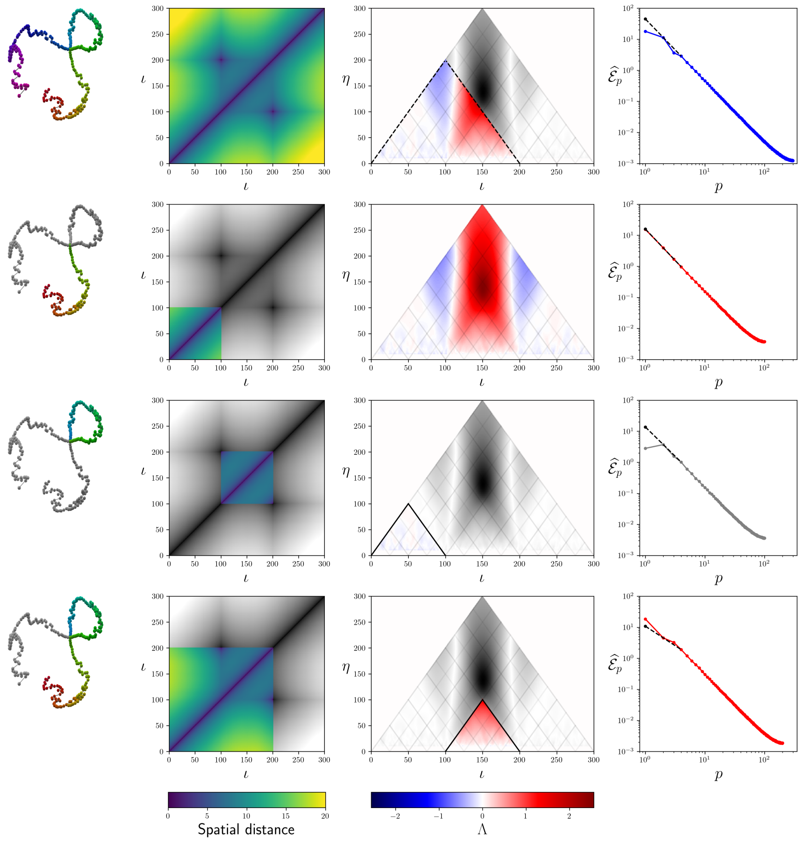


Figure 4.14: Λ-plot for an ensemble of 2000 samples of a (random walk) polymer with $N = 300$ monomers, all containing an internal loop of size 100 in the middle (from index 100 to 200). Different rows focus on distinct sub-regions of the same polymer: whole polymer; first third; inner loop; first two thirds (including the loop). The first column displays a *mean* polymer configuration using a method similar to ShRec3D [48]. Sub-regions are colored accordingly. The second column shows the distance map of the polymer, where coloring focus on the selected region. The third column shows the Λ-plot for this polymer ensemble, with colored triangles highlighting regions corresponding to the selected sub-chain. The spectrum for the selected sub-region is shown in column 4.

coordinate of these maxima precisely corresponds to the midpoint of the loop, while the η coordinate is systematically slightly larger than the actual loop size. Thanks to our straightforward loop modeling, we can derive analytical results, as outlined in the following box.

Maxima coordinates in the Λ-plot

Here I explain how to determine the loop coordinates (ι, θ) from the maxima (ι, η) in the Λ-plot. The midpoint of the loop coincides with the first maxima coordinate and can therefore be directly determined. To relate η at the maximum of the Λ-plot with the actual loop size θ, we perform the following analysis.

We take a cross-section of the Λ-plot for the $\iota = \iota_{\max}$ of the maximum. Any given $\eta < \theta$ corresponds to selecting a sub-walk that is contained inside the loop; for $\eta > \theta$, we are selecting the entire loop and some of the adjacent ends. By modeling this problem as a 3D random walk, we can explicitly calculate the DCT to find the spectrum of both a partial loop ($\eta < \theta$) and a loop with non-looped ends ($\eta > \theta$). From these spectra, we can calculate the log-spectral ratio Λ. By writing $\mu = \eta/\theta$, we can discriminate between the two cases with one parameter and we can go to the continuum limit while keeping the ratio μ fixed. We find then

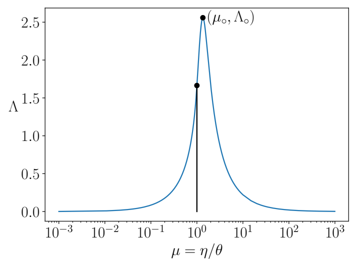
$$\Lambda_0 = \Lambda\big|_{\iota=\iota_{\max}}(\mu) = \begin{cases} -\log\left(1 - \dfrac{8\mu}{\pi^2}\right) & \mu \leq 1, \\ -\log\left(1 - \dfrac{8\mu}{\pi^2}\sin^2\left(\dfrac{\pi}{2\mu}\right)\right) & \mu > 1. \end{cases} \tag{4.8}$$

Note that this function is once differentiable and has a single maximum. In the figure hereafter, we plot equation (4.8).

The condition of zero derivative for the function in equation (4.8) yields $\frac{\pi}{\mu}\sin\left(\frac{\pi}{\mu}\right)+\cos\left(\frac{\pi}{\mu}\right)-1=0$, that is equivalent to solve

$$2x = \tan x \qquad \text{where } \mu = \frac{1}{x}\frac{\pi}{2} \text{ and } x \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right].$$

This equation cannot be solved analytically, but a numerical solution gives $\mu_0 = 1.34767$, with corresponding value $\Lambda\big|_{\iota=\iota_{\max}}(\mu_0) = 2.55882$. Hence, given the position of a maximum (ι, η), the size of the loop is estimated by $\theta = \eta/\mu_0$.



For a given fBm signal containing an internal loop centered at ι_0 with a size of θ, the Λ-plot restricted to the $\iota = \iota_0$ line is indeed given by equation (4.8). This allows a precise determination of the loop position and size starting from the detected maxima (ι, η)[35]. With these results, we can formulate a method for detecting loops in signals. Given a set of signals $\{\vec{x}_n\}$ containing internal loops, follow these steps:

1. Calculate the estimated Λ-plot from the available samples;
2. Find the position $(\iota = \iota_0, \eta)$ of any maximum;
3. Divide η by $\mu_0 \approx 1.34767$ to find the approximate size of the corresponding loop;
4. The estimated loop falls then between monomers $\iota_0 - \eta/(2\mu_0)$ and $\iota_0 + \eta/(2\mu_0)$.

Finally, note that, taken a point (η, ι) on the Λ-plot, the triangle of which it is the vertex corresponds to the lambda plot of the region $[\eta - \iota/2, \eta + \iota/2]$, as shown by the multiple examples given in Figure 4.14.

4.4.4 *Estimating the ratio of looped to non-looped conformations*

In a typical experimental dataset, only a portion of the configurations will exhibit a specific loop, while the complementary fraction will lack this feature. Consequently, we need to investigate how the log-spectral ratio depends on the probability of occurrence of a given loop, and whether it can provide any information about this probability. In the box hereafter we derive an expression for the log-spectral ratio Λ (for fixed $\iota = \iota_0$) for mixed populations in terms of the probability $\mathscr{P}$ of having a loop of size θ. From this expression, we learn that the position of the maximum is independent of $\mathscr{P}$, while its amplitude depends on it. Since we have access to this maximal value of the log-spectral ratio from the Λ-plot, we may use it to know the looping probability $\mathscr{P}$. Indeed, inverting this formula yields

$$p = \frac{\pi^2}{8\mu_0}\left(1 - e^{-\Lambda_{\max}}\right)\csc^2\left(\frac{\pi}{2\mu_0}\right). \tag{4.9}$$

Crucially, this connection between the fraction of looped conformations and the strength of the maxima provides a means to estimate the fraction of looped conformations in the sample, offering valuable insights for biological datasets.

[35] We have $\iota_0 = \iota$ and, from equation (4.8), $\theta = \eta/\mu_0$, where $\mu_0 \approx 1.34767$ is a universal constant.

Log-spectral ratio in mixed populations

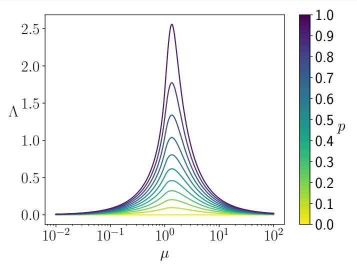
Suppose $\mathscr{P}$ gives the percentage of samples that have a loop in a dataset, and hence that $1-\mathscr{P}$ gives the percentage of samples that do not have a loop. By linearity,

$$\mathscr{P}\left\langle L_p^2\right\rangle + (1-\mathscr{P})\left\langle U_p^2\right\rangle$$

is the spectrum for this mixed population, where the $\left\langle L_p^2\right\rangle$ denotes the spectrum for a uniform population of polymers with the given loop and $\left\langle U_p^2\right\rangle$ denotes the spectrum of a non-looped population. Using this expression within the definition of the log-spectral ratio, we find after going to the continuum limit

$$\Lambda_{\text{mixed}}(\mu) = \begin{cases} -\log\left(1-\dfrac{8}{\pi^2}\mathscr{P}\mu\right) & \mu \le 1, \\ -\log\left(1-\dfrac{8}{\pi^2}\mathscr{P}\mu\sin^2\left(\dfrac{\pi}{2\mu}\right)\right) & \mu > 1. \end{cases} \tag{4.10}$$

The theoretical midlines of Λ-plot for different values of $\mathscr{P}$ are shown in the graph. For $\mathscr{P}=1$, the plot of equation (4.8) is recovered. For $\mathscr{P}=0$, the log spectral ratio vanishes. For other values of $\mathscr{P}$, we plot equation (4.10). The maximum of each function remains at $\mu=\mu_0$, but the value of the maximum decreases as $\mathscr{P}$ decreases.



This equation relies on the intensity of the Λ signal to the probability $\mathscr{P}$. By taking the equation at the observed maximum, $\Lambda_{\text{mixed}}(\mu_0)=\Lambda_{\max}$, and by inverting it, we can therefore recover an estimate of the proportion $\mathscr{P}$ of samples with loops as

$$p = \frac{\pi^2}{8\mu_0}\left(1-e^{-\Lambda_{\max}}\right)\csc^2\left(\frac{\pi}{2\mu_0}\right). \tag{4.11}$$

4.5 Λ-PLOT LOOP DETECTION IN MULTIPLEXED FISH DATA

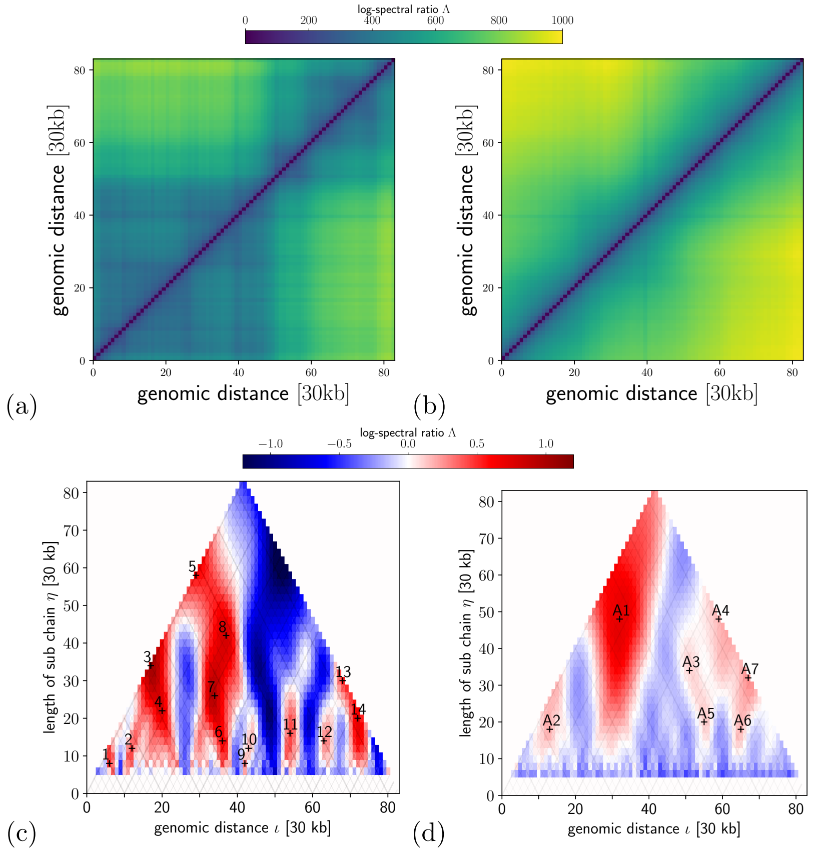
To evaluate the performance of the Λ-plot method on experimental data, we turn again to the **seq-FISH datasets by Bintu et al.**[36]. We consider **experiments 5 and 6** defined in section 4.2 (see Table 4.1), the HCT116 cell line, from 34.6 Mb to 37.1 Mb, with and without RAD21 depletion.

In Figure 4.11, we first present the average distance maps we obtained for the two datasets. In the wild-type data, two large TADs are evident, along with numerous sub-TADs. However, identifying specific loops is challenging. In the auxin-treated variant, the (sub-)TADs are less pronounced, and a significant loss of structural detail is observed at the ensemble average level. No distinct loop can be identified from the distance map.

The log-spectral ratio successfully detects numerous loops in the conformations, including **14 loops in the wild type** (labeled 1 to 14) and **7 loops in the auxin-treated variant** (labeled A1 to A7). Maxima detection involves manual selection of regions where they might be present, followed by standard numerical methods. The estimated proportions of looped conformations for each loop, along with their respective errors, are summarized in Figure 4.16 (red bars).

Examining the relative positions of loops is also interesting. Some

[36] B. Bintu, L. J. Mateo *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, 2018, [9].



(a) (b) (c) (d)

Figure 4.15: The Λ-plots **(a,b)** and distance maps **(c,d)** created from the experimental data of Ref. [9] for both the wild type variant (left) and an Auxin-treated variant (right). Each detected maximum is given a loop id.

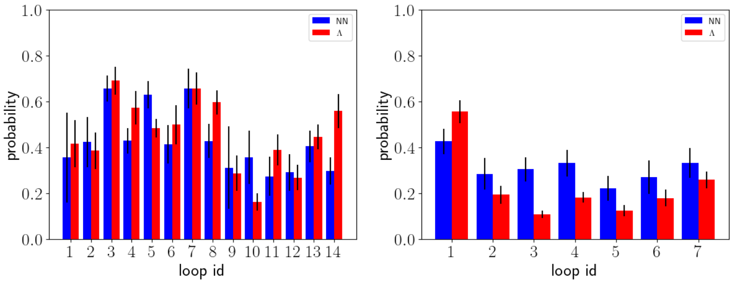


Figure 4.16: The estimated probability of each loop's occurrence is shown, obtained from the Λ-plot (red bars) and the neural network output (blue bars).

loops overlap or are **included within larger loops**, as can be understood by visualizing the corresponding triangles in the *Lambda* diagrams. For example, loops 12 and 13 are inside loop 14, while loop 11 is relatively isolated from the others. In the auxin case, loops A3, A5, A6, and A7 are within loop A4, and loop A1 partially overlaps with loop A2.

Notably, some loops are **closely adjacent to each other**, such as loops A3 and A7 or loops 12 and 14, forming what appears to be the two "petals" of a flower-like shape. It's interesting to note that Ref. [37] suggests that flower-like looping is a fundamental mechanism in chromatin folding leading to hubs or clusters of interacting cis-regulatory modules including enhancers and promoters. This suggests that our algorithm is capable of detecting such structures. However, it's important to confirm that these loops are present simultaneously in unique

[37] S. M. Espinola, M. Götz *et al.* Cis-regulatory chromatin loops arise before tads and gene activation, and are independent of cell fate during early drosophila development. *Nature Genetics*, 53(4):477–486, 2021, [21].

loop id	1	2	3	4	5	6	7
ι_0	6	12	17	20	29	36	34
η	8	12	34	22	58	14	26
Loop range	3-9	8-16	4-30	12-28	7-51	31-41	24-44

loop id	8	9	10	11	12	13	14
ι_0	37	42	43	54	63	68	72
η	42	8	12	16	14	30	20
Loop range	21-53	39-45	39-47	48-60	58-68	57-79	65-79

loop id	A1	A2	A3	A4	A5	A6	A7
ι_0	32	13	51	59	55	65	67
η	48	18	34	48	20	18	32
Loop range	14-50	6-20	38-64	41-77	48-62	58-72	55-79

Table 4.2: The ι_0 and η coordinate of the maximum in the Λ-plots for each loop identified in Figure 4.15. The inferred loop extremities $\iota_0 \pm \eta/(2\mu_0)$ are also listed.

configurations, rather than being a result of averaging across the entire dataset. To address this question, we need to determine **in which specific samples a detected loop is present**. This will be explored in the next subsection.

4.5.1 *Using neural networks to segregate looped and non-looped configurations*

To validate and complement our log-spectral method, Michaël **developed a neural network (NN) approach** to assess the presence of specific loops in individual conformations. The neural network was trained using artificially generated looped and non-looped random walks, employing 20,000 training samples, 5,000 validation samples, and 2,000 test samples in each category[38]. More details are provided in the Appendix, see section 4.7.

To independently gauge the presence of loops, our NN is fed with the spatial distances between pairs of points equidistant from the loop midpoint ι along the chain. In looped configurations, these distances should exhibit a minimum at approximately half the loop's length, while in non-looped ones, on average, they should show a linear increase with distance (see Figure 4.23). Once a loop is identified, by locating a maximum (ι, η) in the Λ-plot of FISH data, we use the trained NN on each individual conformation **to ascertain whether it contains a loop at the specified position**.

The neural network approach offers the added benefit of allowing us to collect data at the single-cell level, enabling further analyses on segregated datasets. We emphasize that the Λ-plot is pivotal for the NN's applicability, since the NN can only be applied to one location at a time and is specifically trained for a single loop size.

As a first test, we compare the proportions of looped conformations determined by the neural networks to those obtained from the Λ-plot for each detected loop in the datasets. The results are presented in Figure 4.16. Hypothesis testing reveals that we can reject the null hypothesis of equal estimated proportions for all loops except for loops

---

[38] Generating our own training data is highly advantageous, as it avoids using experimental data for network training, effectively minimizing data wastage.

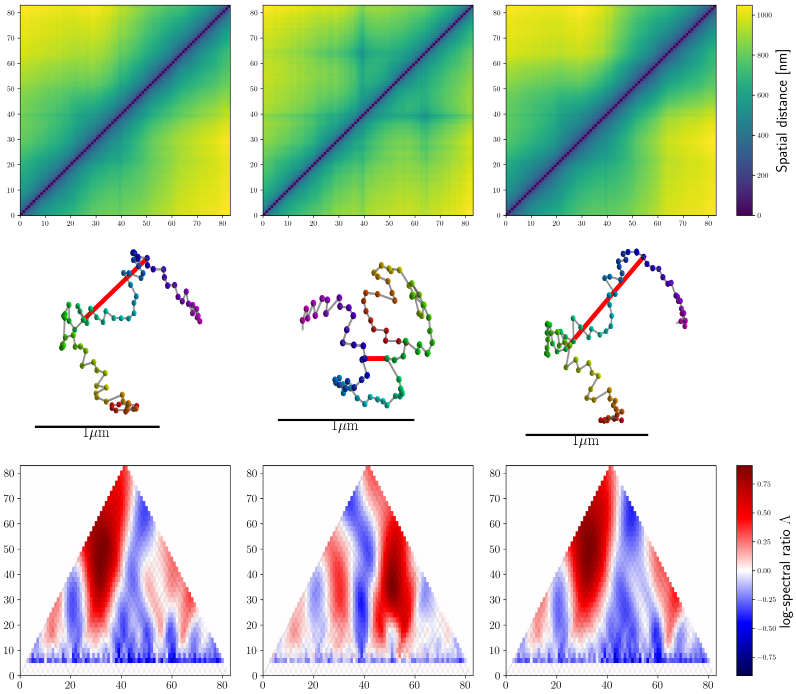


Figure 4.17: Typical example of the output of neural network segregation of looped and non-looped populations, for loop A3. The first column shows the distance map, Λ-plot, and mean configuration (similar to the ShRec3D algorithm [48]) for all measurement data. The second and third columns show the distance map, Λ-plot, and mean configuration for measurements that the NN recognized as containing or lacking loop A3, respectively.

14 and A3 (or 5, 14, A3, and A4) with 99% (or 95%) confidence. This strong agreement between the two methods underlines their reliability.

The NN approach enables precise discrimination **at the individual conformation level** once a loop is detected in the population by the lambda-plot method. This, in turn, enables the separation of two distinct sub-populations: one with looped configurations and the other with non-looped configurations. For illustration, in Figure 4.17, we present a comparison of average distance maps for the whole Auxin-treated dataset and those derived from its sub-populations - one with loop A3 and the other without, as determined by our NN approach.

Strikingly, in the looped sub-population distance map, a local maximum, a typical indicator of loops in contact maps, appears at the position of the predicted loop.

→ *Neural network applied to non-looped regions*

To check the occurrence of false positives in the NN loop detection, we select a random region (from 0kb until 8 × 30kb) which, according to the Λ-plot, contains no loops. Note moreover that this is a small region, which is generally more difficult for the neural network to work with. Figure 4.18 shows the estimated loop-probability (output of the NN) for the selected region and, for comparison, for the region of loop A1. The two outputs are qualitatively different. The distribution of the looped region is bimodal, indicating the presence of a looped sub-population while that of the random region has a single mode. Moreover, the random region displays a steep cut-off before reaching a probability of one being looped.

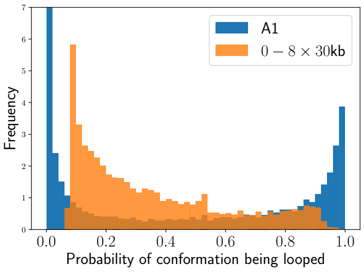


Figure 4.18: Histogram of the probability of a single measurement configuration being looped, for loop region A1 and for the non-looped region 0 − 8 × 30kb.

→ *Reconstructed* average *conformations*

Additionally, we include the corresponding mean configurations, reconstructed following the method outlined in the following box, which provides additional confirmation of the NN's effectiveness in distinguishing configurations containing a loop within the region pinpointed by the Λ-plot approach.

Mean polymer configuration: ShRec3D-like approach

The ShRec3D algorithm [48] is aimed to reconstruct spatial distances and three-dimensional genome structures from observed contacts between genomic loci. In the data from multiplexed super-resolution, the single configurations are known. However, we follow a simplified approach in the spirit of the ShRec3D algorithm in order to have a representation of the *average* features of an entire dataset. To this aim, we calculate individual distance maps for each configuration, then average over all these maps. This average map will be invariant to translations and rotations of each individual polymer. Moreover, the averaged map will still be a distance map (i.e. be symmetric and satisfy the triangle inequalities). Hence, we can choose to put the first monomer in the origin, the second monomer on the positive x-axis, and the third monomer on the $z = 0$-plane, and then the distance map completely determines the polymer configuration. This is then the average configuration.

To further validate the method's accuracy, we implemented the NN procedure on regions identified by the Λ-plots as lacking loops. The results demonstrate the NN's capability to correctly discern the absence of a significant looped sub-population.

By separating the looped and non-looped conformations, we've been able to investigate the relationships between loops, specifically the joint probabilities of each loop pair. In Figure 4.19, we present the Pearson correlations for the loops in the experimental data. All loops in Auxin-treated, except loop A2, are positively correlated with each other. However, loops A3, A4, and A7 seem to be correlated to each other pairwise, consistent with the idea of A3 and A7 forming the two petals of a flower-like shape, where the combination of A3 and A7 is the loop A4. Similarly, on the wild-type variant, loop 13 is the combination of loops 12 and 14. Loops 12 and 13, as well as 13 and 14, are positively correlated, while loops 12 and 14 are anti-correlated. This suggests that the flower-like shape is less likely to occur than the two individual loops separately. Instead, it seems the flower-like shape only emerges from averaging over multiple cells.

In the wild type, it is also remarkable that loop 11 seems to be independent of the other loops. Furthermore, it's interesting to observe anti-correlations between neighboring loops, such as loops 1 and 2, loops 6 and 10, and loops 12 and 14. This might suggest an underlying biological mechanism that prevents adjacent loops from occurring simultaneously.

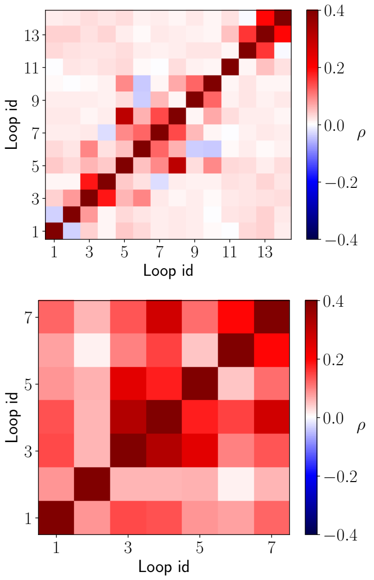


Figure 4.19: Pearson correlation between loops in both wild type (upper plot) and Auxin-treated (lower plot) datasets.

4.5.2 *End-to-end distance distributions differ for looped and non-looped populations*

The distributions of end-to-end distances - the distance between the two extremities of the loop - are shown in Figure 4.20, and reveal variations between looped and non-looped populations in the FISH data for both the wild type and Auxin-treated cases. These populations are considered separately for comparison. In the looped population, a prominent peak at shorter distances is evident, whereas the non-looped population exhibits a broader distribution centered on larger distances and growing with the loop size, in agreement with what is expected for linear polymers.

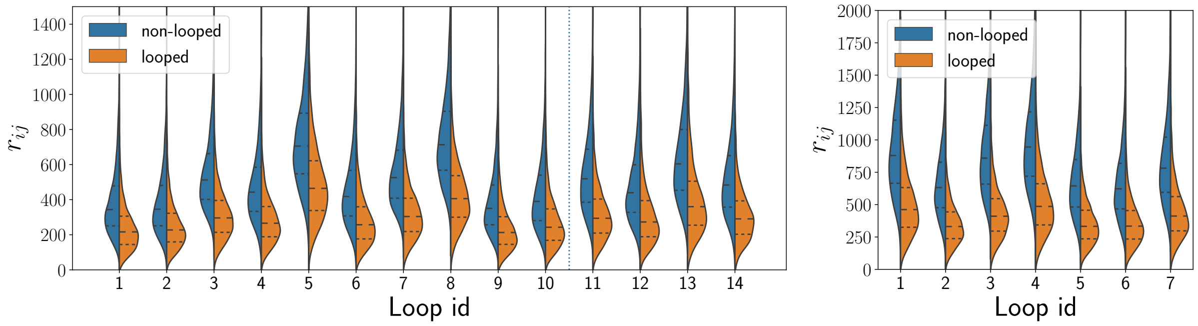


Figure 4.20: Distributions of end-to-end distances r_{ij} which measure the separation between the two extremities $i = \iota_0 - \eta/(2\mu_0)$ and $j = \iota_0 + \eta/(2\mu_0)$ for looped (orange) and non-looped (blue) configurations across all loops identified in the FISH data. Left: wild-type; right: Auxin-treated.

It's important to emphasize that there is a significant overlap in the end-to-end distance distributions between these two populations. This finding demonstrates the inadequacy of a simple analysis of inter-loci distances for loop discrimination and points to the need for a more comprehensive approach, as proposed in this study.

4.5.3 *Further insights into chromatin architecture*

We can use the analogy of fBm to gain further insights into chromatin architecture features in TADs. Let's consider the two large TADs in the wild-type (from 0 to $50 \cdot 30$kb, region (1), and $50 \cdot 30$ kb until $83 \cdot 30$kb, region (2)) and the entire region in the Auxin-treated dataset (Region 3) as a potential third TAD. If we treat these regions as non-looped, we can fit the internal end-to-end distance $R(s)$ with a power law $f(s) = A(s/30\text{kb})^H$, for each of these regions. The fitted exponents H are given in the first row of Table 4.3.

region	(1)	(2)	(3)
all conformations	0.283 ± 0.014	0.314 ± 0.014	0.300 ± 0.003
low loop content	0.394 ± 0.015	0.418 ± 0.016	0.411 ± 0.005

Table 4.3: Values of the exponent H obtained by fitting $R(s) = A(s/30kb)^H$ for regions (1), (2) and (3) (see main text) while considering all the conformations (**upper row**) or only conformations with two loops or less (for the wild type) or without any loops (for the Auxin-treated variant) (**lower row**).

If, now, we only select the subpopulation containing two loops or fewer, for the wild-type, and the subpopulation without any loops, for the Auxin-treated variant, we find different H exponents, as shown in the second row of Table 4.3. **By excluding looped populations, the fitted exponents change** notably, going from **0.3 (looped) to 0.4 (non-lopped)**. Hence, our approach, by allowing us to segregate looped and non-looped populations, enables a more accurate interpretation of the chromatin structure.

Before passing to a general conclusion of the chapter and of the first part of the thesis, here is a schematic resume of the main findings of our loop detection approach.

Spectral loop detection in FISH data

The specific expertise we acquired in the spectral point of view on polymer conformations allows us to approach the question of loop detection from an original point of view.

- The Λ-Plot relies on spectral features to distinguish looped from non-looped chromatin configurations and effectively identifies loops in the ensemble of chromatin configurations.
- A neural network was used for validation and confirms the accuracy and effectiveness of the Λ-Plot in detecting loops in the dataset.
- Through the Λ-Plot and the NN, we characterize the detected loops in terms of their size, position, and prevalence within cell populations.
- Thanks to this new approach, we have identified 14 loops in the MERFISH datasets by Bintu et al., acquired from HCT116 cells of human chromosome 21, and 7 loops in the Auxin-treated variant, whose role certainly will motivate further investigations.

4.6 CONCLUSION/TAKE HOME MESSAGE

In this chapter, which concludes the first part of this thesis, I presented concrete applications of the PSD analysis of polymer conformations defined in chapter 2. I started by analyzing the PSD of human chromatin regions, comprising several TADs. In the variant **where RAD21 is depleted**, inducing the vanishing of TAD structures, the behavior of the PSD was compatible with an interacting, finite-sized polymer of interaction parameter of $\epsilon = 0.3$, placing it close to the coil-globule phase transition. Hence, we can conclude, along with Lesage et al., that "naked" chromatin in physiological conditions behaves as a **critical self-attracting polymer**. In single conformations, this critical state of the polymer is characterized by a partially collapsed state, with **nucleation sites of collapsed subchains** stochastically appearing all across the chain. These structures offer an explanation for the TAD-like patterns in the auxin treated single-cell chromatin conformations observed by Bintu et al. Furthermore, as also pointed out by Lesage et al., a critical state of chromatin is biologically advantageous, since it makes large conformational changes, such as opening and closing chromatin domains, which require little energy to perform. Consequently, small changes in physical conditions can produce large conformational changes. In this regard, we showed that **forming a loop** in a critical chromatin segment was enough **to collapse** the chain within the loop, thereby complementing and reinforcing the hypothesis of TADs formation by stable loop CTCF mediated loops.

Then, in the second part of the chapter, I presented the work done in collaboration with Michaël Liefsoens, which also builds upon the spectral method developed in chapter 2, but this time to **detect loops** in polymer conformation. We found that, through a clever algebraic expression involving the four first modes, we could devise a **spectral observable** extremely effective at detecting loops at the population level. Through this method, we could detect multiple loops in data where the conventional methods of loop detection via distance maps failed to notice their presence.

The presence of loops is confirmed via a neural network approach, which further results in the opportunity to **classify chromatin as looped or non-looped** in each cell. This classification is achieved by assessing the presence of specific loops in each measurement. We have demonstrated the feasibility, speed, and reliability of this process. A significant portion of the success of this neural network approach is attributed to the initial guidance provided by the Λ-plot and the ease with which we can generate artificial training data based on an fBm model.

From the segregated looped conformations, we discovered that the **scaling of** $R(s)$ changed notably between looped and non-looped conformations. The looped conformation presented a more compact conformation than the non-looped, characterized by a scaling exponent of the 0.3 when looped versus 0.4 when open. This further corroborates the picture of chromatin as a polymer at criticality. The associated large fluctuations result in local compaction inhomogeneities that, in turn, explain the more complex spectral behavior observed in the untreated variants.

So we conclude this section, and the first part of the thesis, with a view of chromatin as a structure subject to constant conformational fluctuations, as shown in the animation of Figure 4.7, which can be locally stabilized into a globular form by loop formation, and particularly by the loop extrusion mechanism.

The peculiar motion of the chain in the animation raises the **complementary question of the dynamics of such a chromatin state**. In view of this animation, it is tempting to assume that a highly characteristic dynamic takes place in this critical state.

Hence, in the **next chapters**, we change perspective and investigate the dynamics of polymers through the coil-globule phase transition.

4.7 APPENDIX: NEURAL NETWORK APPROACH

4.7.1 *Generalities*

Neural networks form a class of universal function approximators, meaning that by choosing the appropriate network and giving enough training data, the neural network can in principle mimic any function. In this work, we try to approximate the function that takes a polymer and outputs a yes or no answer to the question 'Does this polymer contain a loop?'. In essence, this means we are applying the techniques of logistic regression on higher dimensional input spaces.

To allow the neural network to mimic the above-specified function, we need to supply it with sufficient training data. This is data where the correct labels (yes or no) are known, so that the neural network can essentially adapt its fitting parameters to better answer the yes or no question. To avoid overfitting, validation data needs to be supplied as well, and separate test data is required to test the accuracy of the model. Since we will be training on (looped) random walks, we can generate as many training; validation; and test samples as needed. This is a crucial benefit of this approach.

The necessity of sufficient training data makes it impossible to start from the measurements directly and just start looking for loops. Indeed, many loops can occur together, or intertwined, and all of these possible configurations need many samples to train on. This is why the Λ-plots developed in this work are used to pinpoint possible locations of loops in the sample which can then, loop by loop, be investigated with a neural network trained to distinguish having one loop or no loops at all.

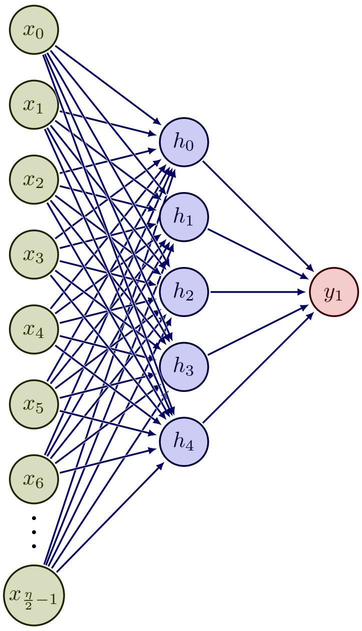


Figure 4.21: The neural network we build to separate looped and non-looped polymers. We input the information of the polymer as sketched in (b top), and the output is a value between 0 and 1 representing the probability that the polymer that was put in is a looped one. The activation function of the hidden layer is the ReLu function, and the sigmoid activation is used for the output layer.

4.7.2 *NN approach for the determination of loops*

The specific networks and inputs we consider, are as follows.

Each time the position of a maximum (ι, η) of the Λ-plot is found, a neural network is trained to separate random walks of length η with and without an internal loop. The loop sizes lie uniformly in the range

$$\frac{\eta}{\mu_0} \pm \max\left(0.1\frac{\eta}{\mu_0}, 1\right),$$

where $\mu_0 = 1.34767$ as previously stated. This range is arbitrarily chosen to give enough variability in the training data so that the neural network can more easily generalize to unseen data. The network has the ReLU-activation function on the hidden layer, and the sigmoid-activation function on the output layer, see for example [39]. We use the binary cross-entropy as the loss function.

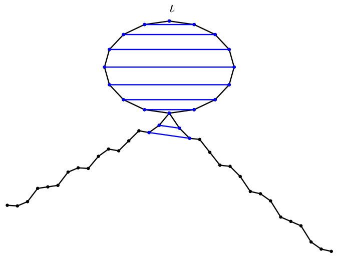


Figure 4.22: As the input of the neural networks, we choose to use the spatial distances between points that are an equal chain distance away from the loop midpoint ι.

The polymer data is inputted as follows. Since the midpoint of the loop is the most certain prediction of the Λ-plot, the distances between the markers and this midpoint are studied. In formulae, we want to represent the (looped) random walk $\{\vec{r}_i\}$ as the signal $\vec{\tilde{x}}$ given by $\tilde{x}_n = \vec{r}_{\iota-n} - \vec{r}_{\iota+n}|$ for $n = 0, \ldots, \eta/2 - 1$. See Figure 4.22 for a sketch of the situation and Figure 4.23 for expected outputs of $\vec{\tilde{x}}$ for

[39] C. C. Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, Cham, 2018, [2].

both looped and non-looped random walks. Since the signal $\vec{\widetilde{x}}$ still has an inherent scale to it, we normalize it to obtain $\vec{x}$ as

$$x_n = \frac{\widetilde{x}_n}{\max(\vec{\widetilde{x}})}.$$

This way, we try to lessen the effect of compact versus loose packing of the chromatin, as well as that of small and large loops.

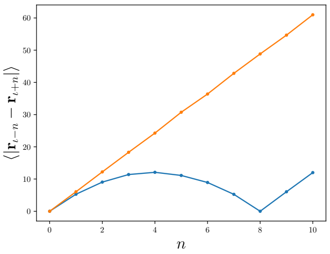


Figure 4.23: We show the expected input for looped (blue) and non-looped (orange) random walks, obtained from averaging over 3000 random walks and 3000 looped random walks.

We train our neural network on artificially generated looped and non-looped random walks—with balanced training data—with 20000 learning samples, 5000 validation samples, and 2000 test samples for both the looped and non-looped random walks. The validation samples are used to monitor and prevent overfitting, and the test samples give an estimate of the accuracy of the neural network. It needs to be remarked that it is an enormous benefit that we can artificially generate training data, as we do not need to waste any experimental data on training the networks.

4.7.3 *Accuracy test*

Additionally, we report the accuracy of each trained network in Table 4.4. Note that this accuracy is obtained by studying independent test data, which is (nevertheless) ideal random walk data, and should hence be interpreted carefully. A conclusion we can make is that for the ideal data, the accuracy is around 90 percent, and that the accuracy is lower for shorter loops. This indicates that the neural networks have more trouble separating short random walks and short looped random walks. This is expected as the thermal fluctuations of each single monomer weigh more heavily on the total conformation, when the total number of monomers is small.

loop id	1	2	3	4	5	6	7
accuracy [%]	81	90	95	95	94	92	92

loop id	8	9	10	11	12	13	14
accuracy [%]	93	83	88	92	92	94	94

loop id	A1	A2	A3	A4	A5	A6	A7
accuracy [%]	95	94	95	95	94	93	94

Table 4.4: Accuracy of trained neural networks for each loop.

4.7.4 *Time complexity*

Creating the Λ-plot requires studying all the sub-polymers at all possible positions, which can be quite time-consuming at first glance. Luckily, due to the application of the Fast Fourier Transform to compute the Discrete Cosine Transform, and by using the fast vectorizing abilities of numerical software like NumPy, this is actually not a problem. Without performing a detailed analysis – since the timing results were satisfactory – we can report that the creation of the two experimental Λ-plots of Figure 4.15 only took about 30 seconds, which is for around 20000 configurations of 83 3D-points each. This timing is for a MacBook Pro with an Apple M1 MAX chip and 32 GB RAM. The training and application of each neural network to each separate loop takes about 11 minutes in total (running in parallel with 10 cores).

PART II : DYNAMICS

INTRODUCTION TO THE SECOND PART

As stated in the introduction, one of the most common methods for probing the physical properties of chromatin is the tracking of a[40] fluorescent segment of DNA over time by optical microscopy. In the videos collected using this method, the segment (a luminous dot) is seen wandering erratically across a black background, with the rest of the DNA chain completely invisible. The trajectory undertaken by the segment is determined by a combination of factors, including its interactions with the particles in the surrounding fluid, the nucleoplasm, as well as its interactions with the rest of the polymer chain. The polymer chain, in turn, is itself erratically evolving through its collisions with the fluid, pushing and pulling the segment[41]. We can therefore reasonably expect that the physical properties of the chain may influence the trajectory of the segment. Consequently, a statistical analysis of these trajectories is frequently employed in the field to deduce certain chromatin properties as chain stiffness, topology, viscoelasticity[42,43,44].

In addition, as shown in the previous part, the folding state of chromatin can be modeled by a self-attracting polymer model. This was shown by studying the **equilibrium properties** of the chain, through the analysis of the average spectral content of the chromatin conformations. The scaling in the equilibrium fluctuations of the low-frequency modes of the chain allowed us to assess the compaction state of the chain. In the same spirit, but in the context of **dynamical data**, one can ask whether it is possible to infer the folding state of a chromatin domain from the trajectory of a single segment belonging to the domain. Or, in the language of polymer physics, **is it possible to infer the position in the coil-globule phase diagram of a polymer from the trajectory of a single monomer?** To answer this question, in this second part of the manuscript, I will attempt to characterize the stochastic dynamics of a monomer within an equilibrium chain in different states of folding[45].

---

[40] Sometimes several, but rarely more than 3 for technical reasons.

[41] And, unfortunately, of many other biological events that tend to make it difficult to interpret the experimental results.

[42] V. I. P. Keizer, S. Grosse-Holz *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605):489–495, 2022

[43] G. Shi, L. Liu *et al.* Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nature Communications*, 9(1):3161, 2018

[44] M. Socol, R. Wang *et al.* Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. *Nucleic Acids Research*, 47(12):6195–6207, May 2019

[45] Note that "dynamics of the coil-globule transition" sometimes refers to the study of the relaxation of a chain whose solvent conditions have been changed abruptly and is therefore out of equilibrium. This question has been extensively studied [32] and will not be the subject of this study.

While the work in this part hasn't been published yet, we are currently in the process of writing an **article incorporating the findings of chapters 6 and 7**.

Also, I would like to mention a very recent article [1] by Amith Z. Abdulla, Maxime M. C. Tortora, Cédric Vaillant and Daniel Jost, in which the impact of topological constraints on the dynamics of a self attracting polymer **embedded** in a non-interacting polymer is investigated. My works and theirs, while certainly overlapping in certain aspects, should be seen as complementary. Indeed, our approach is more focused on the effect of the interacting energy, while their study is primarily concerned with exploring the impact of varying the length of the embedding polymer.

5

Ideal Chain Dynamics: The Rouse model

Our aim here is to qualify the dynamics of a monomer belonging to a self-attractive chain in different solvent conditions. At first sight, this task seems extremely complex, if not completely intractable. Indeed, due to the very strong correlations inherent to polymers as a result of chain connectivity, the dynamics of a monomer depend on the evolution of the entire chain.

In reality, it is, paradoxically, the chain connectivity itself that simplifies the problem. Indeed, the long-range correlations conveyed by the chain connectivity induce fluctuating collective movements of the chain that effectively trump any effect due to the microscopic details of the system. For this reason, the time evolution of polymer conformations is highly universal and can be understood from a few key notions shared by many polymer systems.

Accordingly, the most important model in polymer dynamics is the simplest one, the Rouse model, because it isolates the fundamental property of polymers, namely chain connectivity. Consequently, many more complicated models can be seen as extensions of the Rouse model and share certain properties with it.

Nevertheless, one must be aware that the Rouse model is based on several very rough approximations. Firstly, it's an ideal chain model. As we discussed in chapter 1, in many physical situations monomer interactions are screened and the chain adopts an ideal behavior. We have seen for example in chapter 3 that in globule configurations the polymer subchains behave like ideal chains. For dynamics, we will see that the question is a little trickier, since it won't be possible to apply Rouse dynamics directly to the sub-chains of a globule. Indeed, in this case, neglecting the interaction with the other sub-chains is impossible and gives the wrong scaling for the MSD. To describe this situation, we need therefore to study the dynamics of interactions between sub-chains. To this end, we will introduce the concept of reptation in chapter 7.

Secondly, as we will see, the random force applied to one bead, following the Langevin approach, is completely uncorrelated from the force applied to another bead. This assumption implies that any potential correlations between beads mediated through the solvent are neglected. Other models including the so-called hydrodynamic interactions through random force correlations exist - notably the Zimm model - but won't be treated in this thesis.

In this chapter, in section 5.1, I introduce the Langevin equation, a stochastic differential equation that describes the motion of particles subject to random forces. A particular emphasis will be put on the fluctuation-dissipation theorem, which relates the fluctuating force to an underlying dissipative process.

In section 5.2, we'll apply the Langevin dynamics to a Gaussian chain to derive the Rouse model. To solve the model, we'll again employ the Rouse modes, this time in a more classical fashion. We'll study the dynamical properties of the Rouse modes and see that they describe the motion of the chain on different length scales. Finally, in section 5.3, we'll define the mean square displacement and derive its scaling for a single monomer in a Rouse chain

5.1 The Langevin Approach to Particle Dynamics

Langevin dynamics, which has its origins in the modeling of Brownian motion, is a mathematical formalism that models the trajectory of a large particle exchanging kinetic energy with a heat bath, usually a fluid, composed of smaller particles. The strategy of the Langevin formalism is to model all collisions with the fluid by a random stochastic force acting on the particle, drastically reducing the number of degrees of freedom. The result is an equation similar to Newton's equation but including a stochastic term representing collisions, called the Langevin equation. The Langevin formalism has been generalized to the cases where the "particle" is not an actual particle per se, but rather a collective feature of a macroscopic system. In the following, I recall the main lines of the Langevin approach.

5.1.1 *Underdamped Langevin dynamics and the Langevin equation*

→ *Dissipative process*

Consider a one-dimensional Brownian particle with radius r, mass m, position x, and velocity v, immersed in a fluid with viscosity η. The friction force felt by this bead is proportional to its velocity with proportionality constant given by the Stokes law

$$\zeta = 6\pi\eta r. \tag{5.1}$$

First let's imagine this is the only force acting on the bead, or rather that it dominates the dynamics of the bead. Then Newton's law for this bead gives the equation of motion for the bead:

$$m\frac{dv}{dt} = -\zeta v \tag{5.2}$$

or, by introducing the friction constant $\gamma = \zeta/m$,

$$\frac{dv}{dt} = -\gamma v. \tag{5.3}$$

As a first-order linear differential equation, (5.3) can be solved exactly given the initial velocity v_0 of the bead:

$$v(t) = v_0 e^{-\gamma t}. \tag{5.4}$$

Due to the friction, the velocity of the bead decays exponentially and goes to 0 at long times. This is a typical **dissipative process**. The initial kinetic energy of the bead is transferred, as heat, to the fluid.

→ *The (Underdamped) Langevin Equation for a Free Particle*

This solution is not completely satisfactory because we know that the velocity of the bead should remain non-zero at equilibrium due to thermal fluctuations. In fact, the **equipartition theorem** gives us a precise value for the equilibrium fluctuations of the velocity:

$$\langle v^2 \rangle = \frac{k_B T}{m}. \tag{5.5}$$

Invalidating the assumption that the dissipative force fully dominates the dynamics. The solution put forward by Langevin (in 1908) is to add a random force $\Xi(t)$ acting on the bead. The resulting equation is the simplest possible form of the **Langevin equation**. It describes the dynamics of a free Brownian particle as

$$\frac{dv}{dt} = -\gamma v(t) + \xi(t), \tag{5.6}$$

where $\boldsymbol{\xi(t) = \Xi(t)/m}$ has the units of a force per unit mass, but is often also referred to as the "**stochastic force**", albeit improperly.

Owing to the stochastic nature of the force $\xi(t)$, to complete the Langevin equation we must specify its statistics. Many choices are possible allowing for the modelization of a multitude of random phenomena. In its simplest form, which will concern us in this work, the fluid can be considered isotropic, fixing

$$\langle \xi(t) \rangle = 0. \tag{5.7}$$

Due to the chaotic nature of the fluid-colloid interaction, different collisions are considered independent, i.e

$$\langle \xi(t)\xi(t') \rangle = 2\alpha\ \delta(t-t')\,, \tag{5.8}$$

where $\langle \xi^2(t) \rangle = 2\alpha$, the variance of the random force $\xi(t)$, gives a measure of its intensity.

As we learned in chapter 2, Equations (5.7) and (5.8) define, in practice, a white noise. We may note, then, that in the case where

$\gamma = 0$ the velocity $v(t)$ is simply the primitive of the white noise, and is therefore expected to behave like a Brownian motion, i.e. to have a diverging variance $\langle v^2(t) \rangle$. It is the presence of dissipation (modelled by the term $-\gamma v(t)$) that enables this variance to be locked at the equilibrium value.

→ *Statistics of Solutions to Langevin Equation and the Fluctuation Dissipation Theorem*

After appropriately defining the stochastic force, we can solve formally equation (5.6) for the free particle by integrating both sides from 0 to some arbitrary time t, yielding the general form of the solution,

$$v(t) = v_0 e^{-\gamma t} + \int_0^t dt' \; e^{-\gamma(t-t')} \xi(t), \tag{5.9}$$

where the initial value of the velocity, $v_0 = v(0)$, has been fixed. Due to the presence of the right-hand term, each solution $v(t)$ depends on the specific realization of the stochastic force ξ, meaning that $v(t)$ itself is a stochastic function of time.

Furthermore, we can compute the second moment of the velocity, still at fixed v_0 and therefore denoted $\langle v^2(t) \rangle_{v_0}$, by computing the mean square of (5.9). The two cross terms are first order in the random force ξ and thus vanish. We are left with the two square terms, one of which is a double integral, second order in $\xi(t)$ which can easily be computed by means of equation (5.8), the other is the square of the deterministic exponential term. All in all, we find:

$$\langle v^2(t) \rangle_{v_0} = v_0^2 e^{-2\gamma t} + \frac{\alpha}{\gamma}(1 - e^{-2\gamma t}). \tag{5.10}$$

As anticipated earlier, in the long time limit we expect $\langle v^2(t) \rangle_{v_0}$ to relax towards its equilibrium value $\langle v^2(t) \rangle = k_B T/m$, enforced by the equipartition theorem. Now the $t \gg (2\gamma)^{-1}$ limit of equation (5.10) is α/γ, hence we get

$$\langle v^2 \rangle = \frac{\alpha}{\gamma} = \frac{\langle \xi^2(t) \rangle}{2\gamma} \tag{5.11}$$

$$\boxed{\langle \xi^2(t) \rangle = 2\alpha = \frac{2\gamma k_B T}{m}.} \tag{5.12}$$

This relation fixes the value for the strength of the stochastic force, which was left undetermined. This result is a manifestation of the **fluctuation-dissipation theorem**. In essence, it establishes the relation between the random fluctuations of the force $\xi(t)$, and the strength of the damping term γ such that the process, in its equilibrium state exhibits realistic fluctuations. In hindsight, since both the dissipative and the noise terms result from the interaction of the Brownian particle with the fluid it is not so surprising to find that there is a relation between them.

5.1.2 *Correlation Function and Correlation Time*

From (5.9) we can also compute the **time correlation function** of the particle velocity. First, let's compute the mean velocity at fixed initial velocity:

$$\langle v(t)\rangle_{v_0} = v_0 \; e^{-\gamma t} \tag{5.13}$$

Now we can make the auto-correlation function of the particle velocity appear by multiplying both sides by $v(0) = v_0$ and averaging over initial velocities v_0, yielding

$$C_{vv}(t) = \langle v(t) \; v_0\rangle = \langle v_0^2\rangle \; e^{-\gamma t} = \frac{k_B T}{m} \; e^{-\frac{t}{\tau}} \tag{5.14}$$

where we have defined the **correlation time**[1]

$$\tau = \frac{1}{\gamma} = \frac{m}{\zeta}. \tag{5.15}$$

The parameter τ is the characteristic time of decorrelation of the velocity, i.e. the time needed for the particle to lose memory of its initial velocity v_0. Since τ is also the time taken by the system to fully relax, through the dissipative mechanism, towards its equilibrium state, it is also called the **relaxation time** of the process. If we imagine sampling the signal $v(t)$ with a time step $\Delta t \gg \tau$, then the correlation function (5.14) essentially becomes $(k_B T/m)\; \delta(\tau)$: the sampled velocity is in this regime a delta-correlated, centered stochastic process of variance $(k_B T/m)$ and therefore, again, a white noise. This is of course possible only because the correlation function C_{vv} decreases fast enough - more precisely, it should be integrable.

We can also stress that, although the relation (5.12) is derived in the long-time limit, the typical relaxation time $\tau = \gamma^{-1}$ is in turn linked to fluctuations through the fluctuation-dissipation theorem. Thus, if the solvent's viscosity η is increased at a given temperature, then the velocity relaxes faster (τ decreases), and the amplitude of fluctuations of the random forces due to collisions is greater.

5.1.3 *the Overdamped Langevin Equation*

If we measure the particle's position on long timescales, i.e $t \gg \tau$ we will always miss the inertial regime of the dynamics and measure equilibrated velocities, characterized by $\frac{dv}{dt} = 0$. Therefore we can write the so-called **overdamped** limit of the Langevin equation, by neglecting the inertial term $m\frac{dv}{dt}$:

$$\boxed{\gamma \frac{dx}{dt} = -\frac{dU(x(t))}{dx} + \xi(t).} \tag{5.16}$$

Note that we have also generalized the equation to the case where the particle is subjected, in addition to interaction with the fluid, to a force derived from a potential.

Of course, the solutions to the overdamped Langevin equation are the long time limit of the solutions to the underdamped Langevin equation.

5.1.4 *The Particle Mean Square Displacement*

To measure the spatial extent of a stochastic trajectory, the most common measure is the **mean square displacement (MSD)**, noted $g(t)$,

[1] Note that τ can only be defined when correlations decay rapidly, such as in the case of exponential decay. In the case of a power-law behavior, the correlation function is typically not integrable over time, and this leads to the correlation time diverging.

and defined as the average squared distance covered by the particle in a given time:

$$g(t) = \langle (x(t) - x(0))^2 \rangle. \tag{5.17}$$

By using the relation

$$x(t) - x(0) = \int_0^t dt' \, v(t'), \tag{5.18}$$

the MSD for the Brownian particle can be computed thanks to (5.9) and (5.14). The result is given without going through the details of the calculation:

$$g(t) = 2\frac{k_B T}{m\gamma}(t - \tau + \tau e^{-t/\tau}). \tag{5.19}$$

If we place ourselves at very short time compared to the velocity decorrelation time, $t \ll \tau$, when the particle's inertia dominates its dynamics, we can develop the exponential to second order. The zeroth and first order in the expansion cancel out, leaving only the t^2 term

$$g(t) \sim t^2 \quad | \quad t \ll \tau. \tag{5.20}$$

The t^2 dependence is characteristic of ballistic motion, meaning the particle moves essentially in a straight line in this limit.

In the opposite limit, $t \gg \tau$, we know that $v(t)$ is essentially delta-correlated, as discussed. Consequently, we expect $x(t) - x(0)$ to behave like a Brownian process. Indeed, the exponential in (5.19) goes to 0 and the MSD becomes[2]

$$g(t) = 2\frac{k_B T}{m\gamma}t = 2Dt, \tag{5.21}$$

where we have defined the **diffusion coefficient** of the particle D

$$D = \frac{k_B T}{m\gamma} = \frac{k_B T}{\zeta}. \tag{5.22}$$

Equation (5.21) corresponds to a standard Brownian motion, where the diffusion coefficient D characterizes the average quadratic step size[3]. It is therefore not surprising that D is proportional to the variance of the velocity, $k_B T/m$, as stated by equation (5.22). More interesting, (5.22), called the **Stokes-Einstein relation**, establish a link between D and γ, which is another manifestation of the fluctuation-dissipation theorem.

5.1.5 *Normal, sub- and super-diffision*

It is interesting to derive the position variance $g(t)$ in a more general way, where we only assume that the velocity process $v(t)$ is centered and stationary, of correlation function $C_{vv}(t)$. By following the same procedure as to get the second moment of the velocity (5.10) from the forces, we have to compute the mean square of equation (5.18), which yields

$$g(t) = 2\int_0^t dt' \int_0^{t'} dt'' \, C_{vv}(t''). \tag{5.23}$$

[2] Note that we obtain a square-root dependence on time of the average distance $\sqrt{\langle (x(t) - x(0))^2 \rangle}$, coherently with the behavior evoked in chapter 1 to discuss the internal end-to-end distance (1.13) as a function of the number of steps.

[3] D is also the rate at which this diffusion process occurs at the macroscopic level, process described by the Fick's law.

In the general framework of the under-damped Langevin equation, $C_{vv}(t) = \langle v^2 \rangle \ \exp(-\gamma t)$. We retrieve the same result at large enough time, since, for $t \gg \gamma^{-1}$,

$$\int_0^{t'} dt'' \langle v^2 \rangle e^{(-\gamma t'')} \simeq \int_0^{\infty} dt'' \ \langle v^2 \rangle e^{(-\gamma t'')} = \frac{\langle v^2 \rangle}{\gamma}$$

and therefore

$$g(t) = 2\frac{\langle v^2 \rangle}{\gamma} \ t = 2\langle v^2 \rangle \tau \ t$$

that coincides with equation (5.21). In the case where the overdamped Langevin equation holds, that is, again, for t large enough, $C_{vv}(t)$ can be approximated as a delta-function[4] and we recover the expected behavior even more directly.

This linear dependence of the MSD on time observed for the long-term dynamics of the Brownian particle is typical of so-called **normal** diffusion. It is clear from the previous lines that the linear dependence on time (5.21) emerges whenever the velocity correlation function is integrable, provided that the dynamics is observed on a time-scale much greater than $\int_0^\infty C_{vv}(t)/\langle v^2 \rangle dt$.

If this integrability condition is lacking, diffusive processes exhibit **anomalous diffusion**. Let's consider in particular the case where $C_{vv}(t)$ has a power law behavior. Clearly, the previous derivation fails, and the time-dependence of $g(t)$ becomes non-linear. If the velocity auto-correlation function asymptotically goes as

$$C_{vv}(t) \sim t^{-\beta} \tag{5.24}$$

with β a given exponent in $[0, 2]$, it is simple to show from equation (5.23) that

$$g(t) \sim t^{2H}, \tag{5.25}$$

where $H = 1 - \frac{1}{2}\beta$.

Two cases should then be distinguished:

1. **sub-diffusion**: if $\beta \in [1, 2]$, the particle diffuses less rapidly than in the Brownian case, characterized by $0 \leq H \leq 0.5$;
2. **super-diffusion**: if $\beta \in [0, 1]$, the particle diffuses faster than the Brownian case, i.e. $0.5 \leq H \leq 1$.

Here, we identify H as the Hurst exponent previously introduced. Indeed, different mechanisms can explain the appearance of abnormal diffusion of a stochastic process, most often when the process assumes long-range correlations. A typical case of an anomalous diffusive process is the fBm discussed in chapter 2.

---

[4] Precisely, by using the relation

$$\lim_{a \to 0+} \frac{e^{-|x|/a}}{2a} = \delta(x).$$

5.2 ROUSE DYNAMICS

We can now move on to the study of polymer dynamics. As we shall see, the dynamics of the monomer depend on the evolution of the whole chain and can be understood thanks to a few key concepts relating to the singular evolution of polymer *conformations*.

5.2.1 *The Rouse Equations of Motion: Langevin Dynamics Applied to the Gaussian Chain Model*

The Rouse model describes the motion of a Gaussian polymer, as presented in chapter 1, each of whose monomers is driven by an overdamped Langevin dynamics with uncorrelated white-noise-like forces.

Consider a Gaussian chain polymer of $N+1$ monomers, whose positions are denoted by $(\vec{R}_0, \dots \vec{R}_N)$, immersed in a fluid of viscosity γ at temperature T. As derived in section 1.4, subsequent monomers i and $i+1$ are linked by a bonding potential:

$$U_{i,i+1} = -\frac{1}{2}k(\vec{R}_i - \vec{R}_{i+1})^2 \tag{5.26}$$

where $k = 3k_BT/b^2$, the spring constant, fixes the mean bond length to b according to (1.27). We can then straightforwardly apply the overdamped Langevin equation Equation 5.16 to every bead yielding the system of $N+1$ 3-dimensional **coupled Langevin equations** for the beads

$$\begin{cases} \dfrac{d\vec{R}_0}{dt} &= \dfrac{k}{\gamma}(\vec{R}_1 - \vec{R}_0) + \vec{\xi}_0(t) \\ &\vdots \\ \dfrac{d\vec{R}_i}{dt} &= \dfrac{k}{\gamma}(-\vec{R}_{i-1} + 2\vec{R}_i - \vec{R}_{i+1}) + \vec{\xi}_i(t) \\ &\vdots \\ \dfrac{d\vec{R}_N}{dt} &= \dfrac{k}{\gamma}(\vec{R}_{N-1} - \vec{R}_N) + \vec{\xi}_N(t), \end{cases} \tag{5.27}$$

where we have introduced the random force $\xi_i(t)$ on monomer i with, as seen,

$$\langle \vec{\xi}_i(t) \rangle = 0 \tag{5.28}$$

$$\langle \vec{\xi}_i(t').\vec{\xi}_j(t) \rangle = 6D\delta_{ij}\delta(t-t'). \tag{5.29}$$

with $D = k_BT/\gamma$ the diffusion coefficient (or diffusivity) of a single bead and where δ_{ij} guarantees that the noise felt by one monomer is not correlated to the noise felt by any other monomer, enforcing the **free-draining approximation**.

5.2.2 *Uncoupling the Rouse Equations: The Langevin Equation for the Rouse Modes*

The first step in solving the Rouse model is to write the equations in a matrix form,

$$\begin{pmatrix} \frac{d\vec{R}_0}{dt} \\ \\ \vdots \\ \\ \frac{d\vec{R}_N}{dt} \end{pmatrix} = \frac{k}{\gamma} \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \vec{R}_0 \\ \\ \vdots \\ \\ \vec{R}_N \end{pmatrix} + \begin{pmatrix} \vec{\xi}_0(t) \\ \\ \vdots \\ \\ \vec{\xi}_N(t) \end{pmatrix} \tag{5.30}$$

yielding the first-order differential matrix equation

$$\frac{d\mathbf{R}}{dt} = \frac{k}{\gamma} A\mathbf{R} + \boldsymbol{\xi}. \tag{5.31}$$

In equation (5.31), $\mathbf{R} = (\vec{R}_0 ... \vec{R}_N)^T$ is the vector of position vectors, $\boldsymbol{\xi} = (\vec{\xi_0}, \dots \vec{\xi_N})^T$ the vector of random forces and A is the $N+1$ dimensional matrix appearing in (5.30).

→Equation decoupling

Once written in this form, we immediately recognize that A is the matrix appearing in the Hamiltonian of the Gaussian model (1.31). From there, the strategy is to diagonalize A. We've already done this in chapter 1 to diagonalize the Hamiltonian; it's interesting to find that the **same basis also allows us to decouple the Rouse dynamic equations**, thanks to the simple structure of the Langevin approach.

Let's replace A by its diagonal form (1.45), $A = \Delta^{-1}\Phi D \Phi^T$, in equation (5.31):

$$\frac{d\mathbf{R}}{dt} = -\frac{k}{\gamma}\Delta^{-1}\Phi^T D \Phi \mathbf{R} + \boldsymbol{\xi}. \tag{5.32}$$

To complete the diagonalization of the Rouse equation, remember that ϕ^T is the transition matrix from the real space to the Rouse mode space, in other words $\mathbf{X} = \Phi^T \mathbf{R}$. Hence, we can make the Rouse modes appear by multiplying both sides by Φ, getting

$$\frac{d\mathbf{X}}{dt} = \frac{k}{\gamma} D\mathbf{X} + \boldsymbol{\Xi} \tag{5.33}$$

where we have defined $\boldsymbol{\Xi} = \Phi\boldsymbol{\xi}$.

Finally, by abandoning matrix writing, we can write the **Langevin equations for the Rouse modes**, as

$$\begin{cases} \dfrac{d\vec{X}_0}{dt} & = \vec{\Xi}_0(t) \\ & \vdots \\ \dfrac{d\vec{X}_p}{dt} & = -\dfrac{k}{\gamma}\lambda_p \vec{X}_i + \vec{\Xi}_p(t) \\ & \vdots \\ \dfrac{d\vec{X}_N}{dt} & = -\dfrac{k}{\gamma}\lambda_N \vec{X}_N + \vec{\Xi}_N(t). \end{cases} \tag{5.34}$$

We remind again the formula for the Rouse modes (equation (1.47)),

$$\vec{X}_p = \frac{1}{N+1} \sum_{n=0}^{N} \vec{R}_n \cos\left(\frac{p\pi}{N+1}(n+1/2)\right), \tag{5.35}$$

and the eigenvalues (equation (1.46));

$$\lambda_p = 4\sin^2\left(\frac{p\pi}{2(N+1)}\right). \tag{5.36}$$

→*Statistics of the Random Force Acting on the Rouse Modes*

To complete the Langevin equations of the Rouse modes, we need to specify the statistics of the random forces $\vec{\Xi}_p$. These forces can be explicitly expressed as

$$\vec{\Xi}_p = \frac{1}{N+1} \sum_{i=0}^{N} \vec{\xi}_i \, \cos\left(\frac{p\pi}{N+1}(i+1/2)\right). \tag{5.37}$$

Since $\vec{\Xi}_p$ is a sum of Gaussian random variables, it is itself a Gaussian random variable: Hence, it is sufficient to compute its first two moments.

Clearly, its first moment vanishes:

$$\langle \vec{\Xi}_p(t) \rangle = 0. \tag{5.38}$$

For its second moment, the calculation involves a bit of linear algebra, which is done in the box below, here we simply give the result:

$$\begin{cases} \langle \vec{\Xi}_0(t) \cdot \vec{\Xi}_0(t') \rangle & = \dfrac{6D}{N+1}\delta(t-t') \\ \langle \vec{\Xi}_p(t) \cdot \vec{\Xi}_p(t') \rangle & = \dfrac{3D}{N+1}\delta(t-t') \quad \text{for } p > 0 \\ \langle \vec{\Xi}_p(t) \cdot \vec{\Xi}_q(t') \rangle & = 0 \qquad\qquad\qquad\quad \text{for } p \neq q \end{cases} \tag{5.39}$$

where $D = k_B T/\gamma$ is the single monomer diffusion coefficient.

The dyadic product of two vectors **a** and **b** of size N is an $N \times N$ matrix given by the matrix product of **a** with the transpose of **b** such that its elements are given by

$$(\mathbf{a} \times \mathbf{b})_{ij} = \left(\mathbf{a}\ \mathbf{b}^T\right)_{ij} = a_i b_j \,.$$

Calculating the dyadic product of $\Xi(t)$ with $\Xi(t')$ gives the correlation function between any pair of random forces (Ξ_i, Ξ_j). We have

$$\boldsymbol{\Xi} \times \boldsymbol{\Xi} = \boldsymbol{\Xi}\ \boldsymbol{\Xi}^T = \left(\Phi^{-1}\boldsymbol{\xi}\right)\left(\Phi^{-1}\boldsymbol{\xi}\right)^T,$$

then, by using the fact that $\Phi^{-1} = \Delta^{-1}\Phi^T$ and the identity $(AB)^T = B^T A^T$, we obtain

$$\boldsymbol{\Xi} \times \boldsymbol{\Xi} = \Phi^{-1}\boldsymbol{\xi}\ \boldsymbol{\xi}^T \Phi \Delta^{-1},$$

and taking the ensemble average

$$\langle \boldsymbol{\Xi}(t) \times \boldsymbol{\Xi}(t') \rangle = \Phi^{-1} \langle \boldsymbol{\xi}(t)\ \boldsymbol{\xi}(t')^T \rangle \Phi \Delta^{-1}.$$

Since $\boldsymbol{\xi}(t)\ \boldsymbol{\xi}(t')^T$ is the dyadic product of $\boldsymbol{\xi}(t)$ and $\boldsymbol{\xi}(t')$, we can use the correlation function for ξ and we get

$$\langle \boldsymbol{\xi}(t)\ \boldsymbol{\xi}(t')^T \rangle = 6D\delta(t-t')\underline{I}\,;$$

All in all

$$\langle \boldsymbol{\Xi}(t) \times \boldsymbol{\Xi}(t') \rangle = 2D\Delta^{-1}\delta(t-t'),$$

which amounts to

$$\begin{cases} \langle \vec{\Xi}_0(t) \cdot \vec{\Xi}_0(t') \rangle &= \dfrac{6D}{N+1}\delta(t-t') \\ \langle \vec{\Xi}_p(t) \cdot \vec{\Xi}_p(t') \rangle &= \dfrac{3D}{N+1}\delta(t-t') \quad \text{for } p > 0 \\ \langle \vec{\Xi}_p(t) \cdot \vec{\Xi}_q(t') \rangle &= 0 \qquad\qquad\qquad \text{for } p \neq q. \end{cases}$$

In summary, by diagonalizing the Rouse equations we transformed the $N+1$ coupled equations for the monomers of the chain into $N+1$ decoupled equations for the Rouse modes. We have thus moved **from a local description** of the chain dynamics, i.e. one equation per monomer, **to a global description**, where each equation describes the evolution of a collective motion involving the whole conformation.

5.2.3 *Fluctuation-Dissipation Relation between Rouse Mode and Relaxation Time*

We notice that the Langevin equation for each $p > 0$ Rouse mode, which reads

$$\frac{d\vec{X}_p}{dt} = -\gamma_p \vec{X}_i + \vec{\Xi}_p(t), \tag{5.40}$$

if we define $\gamma_p = \frac{k}{\gamma}\lambda_p$ as the damping coefficient of mode p, retains the same structure as equation (5.6) and thereby describes a **relaxation process** with relaxation time constant

$$\boxed{\tau_p = \gamma_p^{-1} = \frac{\gamma}{4k\sin^2\left(\frac{\pi p}{2(N+1)}\right)} \sim \frac{\gamma}{k}\left(\frac{N+1}{\pi p}\right)^2,} \tag{5.41}$$

where the last equivalence is valid in the $p \ll N$ limit.

Additionally, as seen in chapter 1, the Rouse modes also obey the **equipartition theorem**. Hence we should be able to relate the fluctuations of the random forces $\vec{\Xi}_p$ to the damping coefficient γ_p through the condition of asymptotic equilibrium for the fluctuation of the Rouse mode, identically as we did in the case of pure Brownian motion.

fluctuation-dissipation theorem for the Rouse modes

In that case, the condition was represented by equation (5.11), which we can now directly adapt to the Rouse modes as follows:

$$\boxed{\boxed{\langle X_p^2 \rangle = \tau_p \frac{\langle \Xi_p^2 \rangle}{2}}}. \tag{5.42}$$

Injecting the formulas for τ_p (5.41) and $\langle X_p^2 \rangle$ (1.58) above yields an expression for the random noise,

$$\langle \Xi_p^2 \rangle = 2\frac{3k_BT}{8k(N+1)\sin^2\left(\frac{\pi p}{2(N+1)}\right)} \frac{4k\sin^2\left(\frac{\pi p}{2(N+1)}\right)}{\gamma} = \frac{3D}{(N+1)}, \tag{5.43}$$

which indeed coincides with the result of equation (5.39).

Equation (5.42) is another example of the **fluctuation-dissipa-**

tion theorem. In this case, it relates $\langle X_p^2 \rangle$, which as we have seen in chapter 3, is pertinent to the *structure* of the polymer which is **fluctuation based** in the case of the Rouse model (and polymers in general), to τ_p i.e the relaxation process of the chain or indeed the dynamic evolution of the chain conformation.

5.2.4 *Understanding the mode relaxation times τ_p*

The relaxation times τ_p will play an important role in the determination of the system dynamics. In this respect, several remarks can be made:

1. A typical way to give meaning to these time parameters is to explicitly compute the **correlation function for the Rouse modes**, which in turn will allow us to derive exact formulas for the dynamics of several observables of interest, notably the MSD of the single monomer. Fortunately, we have already solved this problem in the case of the pure Brownian motion. Hence the correlation function[5] of mode p, $C_{pp}(t) = \langle X_p(t) X_p(0) \rangle$ follows the same exponential decay as found in equation (5.14), with time relaxation constant τ_p:

$$\boxed{C_{pp}(t) = \langle X_p^2 \rangle \; e^{-\frac{t}{\tau_p}}}. \tag{5.44}$$

2. The times τ_p also have a clear interpretation from the point of view of the chain fluctuations. In the case of the Brownian particle, the decorrelation time $\tau = 1/\gamma$ is the typical time for the particle to have completely forgotten its initial velocity. Likewise, given an initial conformation $(\vec{X}_0(0), \vec{X}_1(0), \ldots \vec{X}_N(0))$, τ_p is the time after which the $\vec{X}_p(t)$ mode has completely decorrelated from its initial value due to the random fluctuations induced by the heat bath. Now, if we remember the geometrical interpretation of modes discussed in subsection 1.5.1, we can **associate the different $\vec{X}_p$ with different internal motions** of the chain, and the τ_p **with the time needed** to accomplish them. Thus, the decorrelation of mode $p = 1$, which roughly describes the orientation of the chain, essentially involves a rotation of the whole chain and is achieved in a time $\tau_1 \sim N^2$. To decorrelate mode $p = 2$, the chain must perform a bending movement around the middle of the chain, and this takes a time $\tau_2 \sim (N/2)^2$. For the third mode, each half of the chain must perform a bending movement and take an average time of $\tau_3 \sim (N/3)^2$, and so on. Figure 5.1 may help in visualizing this effect.

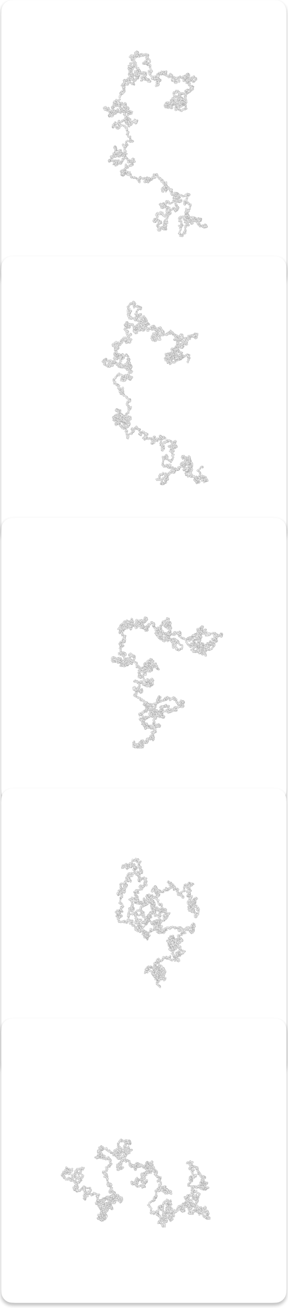


Figure 5.1: By following the temporal evolution of a Rouse chain, we can convince ourselves that larger modes decorrelate over longer times, as discussed in the main text.

3. Importantly, these dynamical considerations also have consequences **on the motion of single monomers**. After a time τ_p, we expect a **monomer to have visited a space, say $g(\tau_p)$, of the extent of a subchain of size** N/p. For a Rouse chain, more precisely:

$$g(\tau_p) \sim b^2 \frac{N}{p}. \tag{5.45}$$

[5] Using the notation adopted in the previous chapter, the correlation of the Rouse modes should be written $C_{X_p X_p}(t)$. I will therefore use the lighter notation C_{pp}, for the aim of readability.

4. In particular, τ_1 is the time needed for a monomer to displace a distance of the order of the **entire chain's extent**, and the longest relaxation time for the system. At the opposite limit, τ_N is the system's shortest decorrelation time, associated with monomer-scale fluctuations. These two extreme correlation times play, as we will see, important roles in the monomer's dynamics, as we will see.

5.3 SINGLE MONOMER DYNAMICS

5.3.1 *Mean Square Displacement: Definitions and Relation to the Rouse Modes*

We can now calculate the mean square displacement, or MSD, of an average monomer. As we shall see, it can be instructive to compute the MSD in different reference frames, namely the lab and center of mass reference frame. Let's start by defining these two observables.

→Single Monomer MSD in the Lab Reference Frame, $g_1(t)$

We define $g_1(t)$, the MSD of a single monomer in a **fixed reference frame**, averaged over all monomers:

$$\boxed{g_1(t) = \frac{1}{N+1}\sum_{n=0}^{N}\left\langle(\vec{R}_n(t) - \vec{R}_n(0))^2\right\rangle} \tag{5.46}$$

→Single Monomer MSD in the Center of Mass Reference Frame, $g_2(t)$

It is also useful to consider the MSD in the **center of mass (CM) reference frame**, $g_2(t)$, i.e the reference frame where the center of mass of the polymer, R_{cm}, defined in (1.15), is fixed at the origin.

Let $\vec{R}_{n,cm} = \vec{R}_n - \vec{R}_{cm}$, be the vector pointing to monomer n in the center of mass reference frame. We define

$$\boxed{g_2(t) = \frac{1}{N+1}\sum_{n=0}^{N}\left\langle(\vec{R}_{n,cm}(t) - \vec{R}_{n,cm}(0))^2\right\rangle.} \tag{5.47}$$

Note that, since $\vec{X}_0 = \vec{R}_{cm}$, in Rouse space, switching to the CM reference frame amounts to setting $\vec{X}_0(t) = 0$ for all t.

→MSD of the Center of Mass, $g_3(t)$

Finally, we define the MSD of the center of mass itself (in the lab reference frame, obviously), whose formula can be expressed in real space or in Rouse space, as

$$\boxed{g_3(t) = \frac{1}{N+1}\sum_{n=0}^{N}\left\langle(\vec{R}_{cm}(t) - \vec{R}_{cm}(0))^2\right\rangle} \tag{5.48}$$

or

$$g_3(t) = \langle(\vec{X}_0(t) - \vec{X}_0(0))^2\rangle, \tag{5.49}$$

by using again the fact that $\vec{X}_0 = \vec{R}_{cm}$.

→*Link Between g_1, g_2 and g_3*

It is possible to express g_1 as a function of g_2 and g_3. Without any loss of generality, we can fix $\vec{R}(0) = 0$, then write

$$g_1(t) = \frac{1}{N+1}\sum_{n=0}^{N}\left\langle R_n^2(t)\right\rangle \tag{5.50}$$

$$= \frac{1}{N+1}\sum_{n=0}^{N}\left\langle (\vec{R}_{n,cm}(t) + \vec{R}_{cm}(t))^2\right\rangle \tag{5.51}$$

$$= \frac{1}{N+1}\sum_{n=0}^{N}\left\langle \vec{R}_{n,cm}^2(t)\right\rangle + \frac{1}{N+1}\sum_{n=0}^{N}\left\langle \vec{R}_{cm}^2(t)\right\rangle + \frac{2}{N+1}\sum_{n=0}^{N}\left\langle \vec{R}_{n,cm}\cdot\vec{R}_{cm}\right\rangle$$

In the last expression, we recognized in the first and second terms, respectively, g_2 and g_3. Furthermore, we have $\left\langle R_{n,cm}\cdot\vec{R}_{cm}\right\rangle = 0$, since moving the CM of the chain has no influence on the position of a monomer in the CM reference frame. This sets the right term to 0, yielding:

$$\boxed{g_1(t) = g_2(t) + g_3(t)} \tag{5.52}$$

which translates the fact that the motion of the monomer can always be decomposed as its motion in the CM reference frame plus the motion of the center of mass.

5.3.2 *Writing $g_1(t)$ in Terms of the Mode Correlations $C_{pp}(t)$*

Fortunately, there exists a simple equation relating $g_1(t)$ to the correlation function of the Rouse modes, $C_{pp}(t)$. Indeed, by using equation (1.48)

$$\vec{R}_n = \vec{X}_0 + 2\sum_{p=1}^{N}\vec{X}_p\cos\left(\frac{p\pi}{N+1}(n+\tfrac{1}{2})\right)$$

to write $\vec{R}_n$ in terms of the Rouse modes, and after some rather tedious algebra, we can show the following, very useful formula:

$$\boxed{g_2(t) = 2R_g^2 - 4\sum_{p=1}^{N}C_{pp}(t),} \tag{5.53}$$

and therefore, equivalently,

$$\boxed{g_1(t) = \underbrace{2R_g^2 - 4\sum_{p=1}^{N}C_{pp}(t)}_{g_2(t)} + g_3(t).} \tag{5.54}$$

Note that **this formula is completely general**, as it only relies on the Rouse decomposition of the position vector $\vec{R}_n$ and **not** on the vanishing cross-correlation of the Rouse modes, as seen in some books. The detailed derivation for (5.54) is given in the following box.

Let's start with the square displacement of monomer n,

$$g_{1,n}(t) = \langle (\vec{R}_n(t) - \vec{R}_n(0))^2 \rangle. \tag{5.55}$$

Now remember that we can always express the position of a monomer in the Rouse coordinates by using equation (1.48) (see just above). We can then inject 1.48 into 5.55, getting

$$g_{1,n}(t) = \left\langle \left(X_0(t) + 2\sum_{p=1}^{N} \vec{X}_p(t) \cos\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right] - X_0(0) + 2\sum_{p=1}^{N} \vec{X}_p(0)\cos\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right]\right)^2 \right\rangle. \tag{5.56}$$

By switching to the CM reference frame by setting $\vec{X}_0(t) = 0$, then factoring the cosines in the remaining terms, we get:

$$g_{2,n}(t) = \left\langle \left(2\sum_{p=1}^{N} (\vec{X}_p(t) - \vec{X}_p(0)) \cos\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right]\right)^2 \right\rangle. \tag{5.57}$$

We can expand the square into diagonal and off-diagonal terms, yielding

$$\begin{aligned} g_{2,n}(t) = & \overbrace{4\left\langle \sum_{p=1}^{N} (\vec{X}_p(t) - \vec{X}_p(0))^2 \cos^2\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right]\right\rangle}^{S^n_{\text{diag}}} \\ & \underbrace{+ 2\cdot 4\left\langle \sum_{p=1}^{N}\sum_{q=1}^{p-1} (\vec{X}_p(t) - \vec{X}_p(0))\cdot(\vec{X}_q(t) - \vec{X}_q(0)) \cos\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right] \cos\left[\frac{q\pi}{N+1}(n+\frac{1}{2})\right]\right\rangle}_{S^n_{\text{off}}}. \end{aligned}$$

Let's start by treating the diagonal term by expanding the square:

$$\begin{aligned} S^n_{\text{diag}} = & 4\sum_{p=1}^{N} \langle X_p^2(t)\rangle \cos^2\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right] + 4\sum_{p=1}^{N} \langle X_p^2(0)\rangle \cos^2\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right] \\ & + 8\sum_{p=1}^{N} \langle \vec{X}_p(0).\vec{X}_p(t)\rangle \cos^2\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right]; \end{aligned}$$

If we assume that $\vec{X}_p(t)$ is **stationary**, we have $\langle X_p^2(t)\rangle = \langle X_p^2(0)\rangle \; \forall t$, meaning the first two terms in S_{diag} are equal. To make further progress we have to **average over all monomers**. By noticing that

$$\frac{1}{N+1}\sum_{n=0}^{N} \cos^2\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right] = \frac{1}{2},$$

we are left with

$$S_{\text{diag}}(t) = \frac{1}{N}\sum_{n=0}^{N} S^n_{\text{diag}} = 4\cdot\frac{1}{2}\cdot 2\cdot \sum_{p=1}^{N}\langle X_p^2(0)\rangle + 8\cdot\frac{1}{2}\sum_{p=1}^{N}\langle \vec{X}_p(0).\vec{X}_p(t)\rangle,$$

$$\boxed{S_{\text{diag}}(t) = 4\sum_{p=1}^{N}\langle X_p^2\rangle + 4\sum_{p=1}^{N} C_{pp}(t),}$$

where $\langle X_p^2\rangle$ is the average PSD and $C_{pp}(t)$ is the auto-correlation function of mode p.

Now we turn our attention to the off-diagonal terms, again averaged over all the monomers:

$$S_{\text{off}} = \frac{1}{N+1}\sum_{n=0}^{N} S^n_{\text{off}} = 8\sum_{p=1}^{N}\sum_{q=1}^{p-1} \langle (\vec{X}_p(t) - \vec{X}_p(0))\cdot(\vec{X}_q(t) - \vec{X}_q(0))\rangle \frac{1}{N+1}\sum_{n=0}^{N} \cos\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right]\cos\left[\frac{q\pi}{N+1}(n+\frac{1}{2})\right].$$

Using the fact that

$$\sum_{n=0}^{N} \cos\left[\frac{p\pi}{N+1}(n+\frac{1}{2})\right]\cos\left[\frac{q\pi}{N+1}(n+\frac{1}{2})\right] = \frac{1}{2}\left[\frac{\sin(p+q)\pi}{\sin(\frac{(p+q)\pi}{2(N+1)})} + \frac{\sin(p-q)\pi}{\sin(\frac{(p-q)\pi}{2(N+1)})}\right] = 0 \quad \text{for } p \neq q$$

we finally obtain $\boxed{S_{\text{off}} = 0.}$ All in all, since $\sum_{p=1}^{N}\langle X_p^2\rangle = R_g^2$, we get the expected result of equation (5.54), under the form:

$$g_2(t) = g_1(t) - g_3(t) = 2R_g^2 - 4\sum_{p=1}^{N} C_{pp}(t).$$

5.3.3 *Monomer Dynamics in the Rouse Model*

Given the previous, general formulas, we can now calculate $g_1(t)$, $g_2(t)$, and $g_3(t)$ for the specific case of the Rouse chain.

→ *The Motion of the Center of Mass: $g_3(t)$*

It is easy to derive $g_3(t)$ for the Rouse chain directly from the Langevin equation for the 0th mode, $\vec{X}_0$. The Langevin equation for $\vec{X}_0$ is indeed:

$$\frac{d\vec{X}_0}{dt} = \vec{\Xi}_0(t) \tag{5.58}$$

with

$$\langle \vec{\Xi}_0(t) \cdot \vec{\Xi}_0(t') \rangle = \frac{6D}{N+1}\delta(t-t'). \tag{5.59}$$

We recognize the overdamped Langevin equation for a 3D free Brownian particle. The associated diffusion coefficient, denoted as D_{cm}, is expressed as:

$$D_{cm} = \frac{D}{N+1} = \frac{k_B T}{m(N+1)\gamma} \quad \sim N^{-1}, \tag{5.60}$$

so that we can write

$$\boxed{g_3(t) = 6D_{cm}t \quad | \quad D_{cm} = \frac{D}{N+1}.} \tag{5.61}$$

This result implies that the center of mass diffusion, which represents the diffusion of the entire chain, behaves like a Brownian particle with mass $(N+1)m$, i.e. the total mass of the chain. Remarkably, this behavior is independent of the chain's shape or folding state. This arises from the free draining approximation (discussed in subsection 5.2.1), which implies that the random net total force felt by the CM is equal to the sum of all forces felt by the monomer.

5.3.4 *Single Monomer Dynamics: $g_1(t)$ and $g_2(t)$*

Moving on to single monomer dynamics, let's compute $g_2(t)$, from which we can deduce $g_1(t)$ using equation (5.53).

By writing $R_g^2 = 2\sum_{p=1}^{N}\langle X_p^2\rangle$ (from Parseval's theorem), we can factor $\langle X_p^2\rangle$, which yields

$$\boxed{g_2(t) = 4\sum_{p=1}^{N}\langle X_p^2\rangle \left(1 - \ e^{-\frac{t}{\tau_p}}\right).} \tag{5.62}$$

To make the computation more readable, let's write, by using equation (5.41) with $N \approx N+1$,

$$\tau_p = \tau_N \left(\frac{N}{p}\right)^2,$$

and, by using again the fluctuation-dissipation relation,

$$\langle X_p^2\rangle = \frac{3D}{2(N+1)}\tau_p = \frac{3D}{2(N+1)}\tau_N \left(\frac{N}{p}\right)^2.$$

Equation (5.62) then becomes

$$g_2(t) = 6D\tau_N \left[\sum_{p=1}^{N}\frac{1}{N}\left(\frac{N}{p}\right)^2 \left(1 - \ e^{-\frac{t}{\tau_N}\left(\frac{p}{N}\right)^2}\right)\right]. \tag{5.63}$$

We may now note that, if we define a new variable $q = p/N$, then q partitions the real interval $[0, 1]$ with step $\Delta q = \frac{1}{N}$, leading to

$$g_2(t) = 6D\tau_N \left[\sum_{q=1/N}^{1} \Delta q \frac{1 - e^{-q^2 \frac{t}{\tau_N}}}{q^2} \right] ; \tag{5.64}$$

Inside the brackets, we recognize the Riemann sum for the following integral:

$$I(t) = \int_{1/N}^{1} dq \; \frac{1 - e^{-q^2 \frac{t}{\tau_N}}}{q^2} . \tag{5.65}$$

The integral of equation (5.65) converges and was computed with the help of `mathematica`, leading to the following approximate expression for $g_2(t)$:

$$\begin{aligned} g_2(t) = 6D\tau_N \Bigg(& \left[\sqrt{\pi \frac{t}{\tau_N}} \operatorname{erf}\left(\sqrt{\frac{t}{\tau_N}} \right) \; + \; e^{-\frac{t}{\tau_N}} - 1 \right] + \\ & + N \left[\sqrt{\pi \frac{t}{\tau_1}} \left(-\operatorname{erf}\left(\sqrt{\frac{t}{\tau_1}} \right) \right) \; + \; 1 - e^{-\frac{t}{\tau_1}} \right] \Bigg) \end{aligned} \tag{5.66}$$

where $\operatorname{erf}(x)$ is the error function[6]. Note that the only assumption made is that N is sufficiently large for the sum to converge to the integral, hence this expression describes the MSD in the CM reference frame **for all** t.

Clearly, the two relevant time scales that appear in $g_2(t)$ **are** $\tau_N \sim \gamma/k\pi^2$ **and** $\tau_1 = (N+1)^2 \tau_N$, the shorter and longer mode relaxation times. Their presence leads to **3 distinct scaling regimes** for $g_1(t)$:

1. A **short** time regime, where t is much shorter than the shortest relaxation time: $t \ll \tau_N$;

2. A a **long** time regime, where t in much greater than the longest relaxation time of the polymer: $t \gg \tau_1$;

3. And an **intermediate** time scales $\tau_N \gg t \gg \tau_1$.

In the following, I'll describe the behavior of $g_2(t)$ in each regime.

→*Cross-over functions*

For the discussion that follows, it's useful to give names to the two following limit functions of $g_2(t)$:

$$g_2^{t \ll \tau_1}(t) = \lim_{t \ll \tau_1} g_2(t) = 6D\tau_N \left[\sqrt{\frac{\pi t}{\tau_N}} \operatorname{erf}\left(\sqrt{\frac{t}{\tau_N}} \right) \; + \; e^{-\frac{t}{\tau_N}} - 1 \right]$$

and

$$g_2^{t \gg \tau_N}(t) = \lim_{t \gg \tau_N} g_2(t) = 6D\tau_N N \left[\sqrt{\frac{\pi t}{\tau_1}} \left(1 - \operatorname{erf}\left(\sqrt{\frac{t}{\tau_1}} \right) \right) + 1 - e^{-\frac{t}{\tau_1}} \right] .$$

Hence, $g_2^{t \ll \tau_1}$ describes the cross-over between the short and intermediate times scale while $g_2^{t \gg \tau_N}$ between the intermediate and long timescale.

[6] The error function $\operatorname{erf}(x)$ is a sigmoid special function defined as:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz.$$

→*Short time regime:* $\boldsymbol{t < \tau_N}$

Let's start by computing the time scaling law for $g_2(t)$ in the short time regime from the cross-over function $g_2^{t\ll\tau_1}$. If t is much shorter than the shortest relaxation time of the polymer, i.e., when $t < \tau_N$, we can expand both the exponential and the erf function to the first order in t/τ_N, to get

$$g_2(t) \simeq g_2^{t\ll\tau_1} \simeq 6D\tau_N \left[\sqrt{\pi\frac{t}{\tau_N}}\frac{2}{\sqrt{\pi}}\sqrt{\frac{t}{\tau_N}} + (1-\frac{t}{\tau_N}) - 1\right]$$

or

$$g_2(t) \simeq 6Dt \quad \text{for } t < \tau_N \tag{5.67}$$

which is the diffusive behavior expected for a **single unconstrained diffusing monomer**, outside of a chain. This outcome can be readily understood: in the short time limit, the monomer doesn't feel the springs linking it to its neighbors, and therefore diffuses freely.

We can also retrieve $g_1(t)$ from equation (5.52), yielding

$$\boxed{g_1(t) = 6D_{cm}t + 6Dt} \quad \text{for } t < \tau_N \tag{5.68}$$

where we used equation (5.61) to express $g_3(t) = 6D_{cm}t$.

Finally, remembering that $D_{cm} = D/(N+1)$, we can conclude that for a long enough polymer $6D_{cm}t \ll 6Dt$. Therefore, in the short time limit, we find

$$g_1(t) \sim g_2(t) \sim 6Dt \quad \text{for } t < \tau_N. \tag{5.69}$$

→*The Intermediate or* ***Rouse regime:*** $\boldsymbol{\tau_N \ll t \ll \tau_1}$

The interesting behavior happens at intermediate times in the so-called **Rouse regime** where the **internal dynamics** of the chain, i.e. the relaxation of the Rouse modes, dominate the dynamics of the monomer. Lets compute its behavior by assuming first $t \ll \tau_1$, meaning $g_2(t) \simeq g_2^{t\ll\tau_1}$, and then $t \gg \tau_N$, which send the exponential to 0 and the erf function to 1, yielding:

$$g_2(t) = 6D\tau_N\sqrt{\pi\frac{t}{\tau_N}} - 1; \tag{5.70}$$

since $t \gg \tau_N$, then $t/\tau_N \gg 1$, leaving us with:

$$\boxed{g_2(t) \simeq 6D\sqrt{\pi\tau_N}\, t^{1/2} \quad \text{for } \tau_N \ll t \ll \tau_1} \tag{5.71}$$

meaning that, in this intermediate regime, the monomer experiences **sub-diffusive anomalous dynamics** as a consequence of its being part of the chain.

→*The long time regime:* $t \gg \tau_1$

Unfortunately, our approximate formula (5.66)for $g_2(t)$ breaks down at the cross-over between the intermediate and long time regime. Indeed, if the limit $t \gg \tau_1$ is taken in $g_2^{t\gg\tau_N}(t)$, the function is found to converge to the wrong constant (see Figure 5.2). Nevertheless, we can easily compute the long-time dynamics of the monomer by taking

$t \gg \tau_1$ in equation (5.62). All the exponentials go to zero in this case, leaving us with

$$\boxed{g_2(t) = 2R_g^2} \quad \text{for } t > \tau_1 : \tag{5.72}$$

in the CM reference frame, the diffusion of a single monomer plateaus[7] at $2R_g^2$.

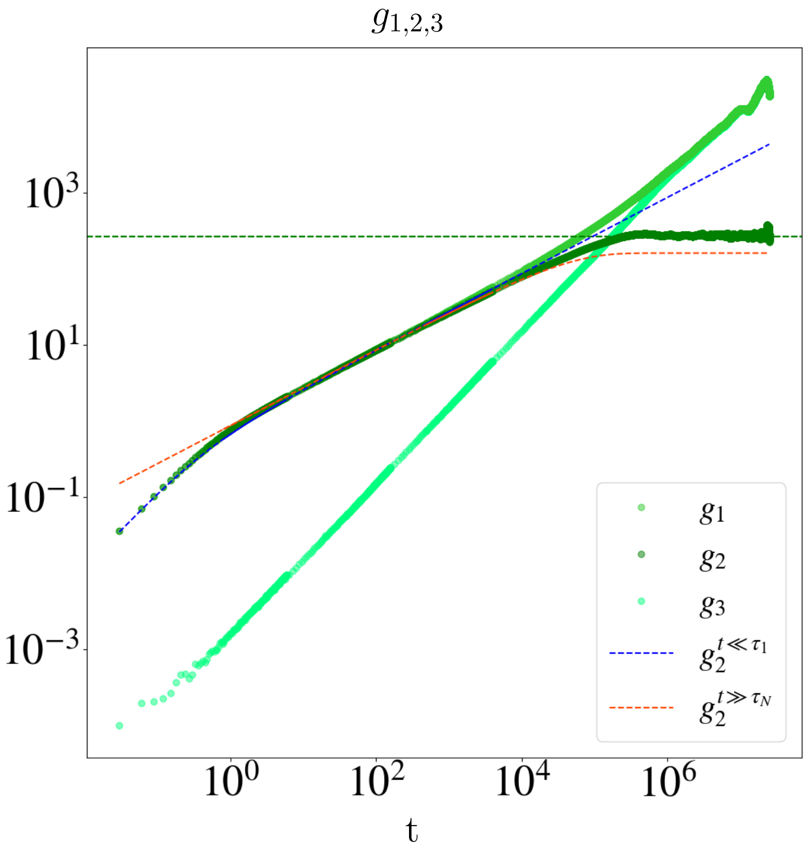


Figure 5.2: g_1, g_2 and g_3 for a Rouse polymer of length $N = 800$ across all three diffusion regimes. The dashed blue are respectively the short and long times approximate cross-over functions $g_2^{t \ll \tau_1}$ and $g_2^{t \ll \tau_N}$. Note how $g_2^{t \gg \tau_N}$ doesn't converge to the correct asymptotic value, $2R_g$, represented by the green horizontal line.

On the other hand,

$$g_1(t) = g_3(t) + g_2(t) = 6D_{cm}t + 2R_g^2 \tag{5.73}$$

give

$$\boxed{g_1(t) \approx g_3(t) = 6D_{cm}t} \quad \text{for } t > \tau_1, \tag{5.74}$$

meaning that **at long times, the motion of an individual monomer reflects the diffusive motion of the entire polymer chain**.

Let's summarize the three different regimes:

Different regimes in the Rouse dynamics:

1. A **short time** regime, where t is much shorter than the shortest relaxation time: $t \ll \tau_N$. In this case the bead experiences **normal diffusion**, as if it wasn't connected to the chain:

$$g_1(t) \sim g_2(t) = 6Dt.$$

---

[7] This result is obviously expected, since monomers can't move any further than the polymer itself.

2. Most interestingly, at intermediate time scales $\tau_N \ll t \ll \tau_1$, the so called **Rouse regime**, where the monomer experiences **subdiffusion with exponent** 1/2:

$$g_1(t) \sim g_2(t) \simeq 6D\sqrt{\pi\tau_N}\; t^{1/2};$$

   As discussed, this is a direct effect of the harmonic connectivity in the chain, that hinders the free diffusion of monomers.

3. Finally, a **long time** regime, where t is much greater than the longest relaxation time of the polymer: $t \gg \tau_1$. In this regime we have

$$g_1(t) \sim g_3(t) = 6D_{cm}t :$$

   the monomer **diffuses coherently with the center of mass** of the chain.
   We also have, in the same regime, that $g_2(t) = 2R_g^2$, meaning that, in the CM reference frame, the MSD plateaus to $2R_g^2$.

5.4 CONCLUSION/TAKE HOME MESSAGES

In this chapter, we began by introducing Langevin dynamics in preparation for the definition of the Rouse model (section 5.1).

Using the example of Brownian motion, we introduced the concepts of correlation function, that is in this case a decreasing exponential function of time, and correlation time of the particle's velocity, capturing the typical time scale that characterizes the particles dynamics.

We then presented the overdamped version of the Langevin equation and carried on with the introduction of another dynamical observable, the mean square displacement (MSD) $g(t)$. In the case of Brownian diffusion, we have shown that the dynamics are divided into a short-time ballistic regime and a linear scaling regime, characteristic of the long-time diffusion, $g(t) = 6Dt$. Importantly, we saw that the characteristic time separating these two regimes is given by the particle velocity correlation time.

Moreover, we established the expression of the diffusion coefficient, $D = k_bT/m\gamma$ using the fluctuation-dissipation theorem, linking the fluid viscosity to the random force fluctuations through the diffusion coefficient via Einstein's law.

To conclude the introductory part of the chapter, we introduced stochastic processes whose increments display power law correlations, inducing a non-linear MSD whit two possible cases: sub-diffusion, if increments are anti-correlated, implying $g(t) \sim t^{2\alpha}$, $\alpha < 0.5$; and super-diffusion: increments are positively correlated, implying $g(t) \sim t^{2\alpha}$, $\alpha > 0.5$.

Thanks to these theoretical tools, we were able to straightforwardly tackle the **Rouse model dynamics** (section 5.2), by applying Langevin dynamics to each bead of a Gaussian polymer. The result is a system of $N + 1$ coupled equations, one per bead, which can be diagonalized by introducing the same Rouse modes already defined in chapter 1. This transforms the equations pertaining to the beads into an equivalent set of equations concerning the Rouse modes. **In the mode representation, the equations are decoupled** and easy to solve. A crucial step in the solution is to calculate the **mode correlation functions** (equation (5.44))

$$C_{pp}(t) = \langle X_p^2 \rangle \, e^{-\frac{t}{\tau_p}} .$$

$C_{pp}(t)$ gives access to the dynamics of each mode, allowing us to apprehend the complex dynamics of the entire chain. Each mode is associated with a characteristic **decorrelation time** (equation (5.41)):

$$\tau_p = \gamma_p^{-1} = \frac{\gamma}{4k \sin^2\left(\frac{\pi p}{2(N+1)}\right)} \underset{p \ll N}{\sim} \frac{\gamma}{k} \frac{(N+1)^2}{\pi^2 p^2} .$$

It's important to have a clear understanding of the role - in fact the roles - of the characteristic times τ_p, as this will enable me to interpret many simulation results correctly. I have proposed **different physical interpretations of this characteristic time**: as the time required to execute a collective movement of the chain involving N/p monomers in unison; as the decorrelation time of a sub-chain of size N/p; as the time needed for a monomer to visit a space whose typical size is the spatial extent of a sub-chain of linear length N/p, i.e $b\sqrt{N/p}$.

Finally, I examined the **MSD of one monomer** in the chain in section 5.3. I started by introducing $g_1(t)$ and $g_2(t)$ the MSDs of the monomer in the lab and center of mass reference frames, and $g_3(t)$, the MSD of the CM of the whole chain, where $g_1(t) = g_2(t) + g_3(t)$.

Thanks to the mode decomposition, I then calculated a **general formula for the MSD** of an average monomer relying on the **correlation function of the Rouse modes** (equation (5.54)):

$$g_1(t) = g_3(t) + 2R_g^2 - 4\sum_{p=1}^{N} C_{pp}(t).$$

With the help of the latter relation, we saw that, in the case of the Rouse model, **the dynamics is partitioned into three regimes, split by τ_1 and τ_N**: A short time regime, where the bead experience normal diffusion; The intermediate **Rouse regime**, where the monomer experiences subdiffusion with exponent 1/2; And a long time regime where $g_1(t) \sim g_3(t) = 6D_{cm}t$ and $g_2(t) = 2R_g^2$.

With this foundation in (ideal) polymer dynamics, we can proceed to analyze the dynamics of collapsing polymers in the subsequent chapter, where we introduce a simplified model for **interacting polymers**.

6

Dynamics in a Collapsing Polymer Without Topological Constraints

In the previous chapter, we studied the dynamics of the Rouse model, an isolated chain whose monomers interact only through chain connectivity. As an ideal chain model, the Rouse model neglects the two main properties of real chains: the **excluded volume** taken up by each monomer and the **attraction** between monomers. Adding solely these properties to an ideal chain leads to models that undergo a coil-globule phase transition[1], as seen in chapter 3 where we studied the **thermodynamics** of the transition. In this section however, we aim to study the dynamics of these polymers.

In the study of real chain dynamics, another property, that does not influence the equilibrium properties of the chain, must be enforced: the chain should be **topologically constrained**. Essentially, what this means is that the chain shouldn't be able to cross itself. To understand the appellation *topologically constrained*, imagine the two ends of the chain to be glued together so that the chain forms a loop. If the chain is not allowed to cross, then the topology of the knot formed by the loop is invariant, i.e. constrained.

As we shall eventually see, in cases where collisions between monomers are frequent, this constraint has a strong influence on the polymer dynamics, complicating substantially the analysis compared to the Rouse chain. This is particularly the case in the highly-studied dense polymer melt. To address this issue, de Gennes has developed a theoretical framework, based upon the idea of reptation, that correctly describes the dynamics of these chains[2].

Most of the time, the excluded volume interaction is used both to model the finite volume of the beads and to ensure the topologically constrained state of the chain, by enforcing that the maximum distance between two adjacent beads is never greater than the size of a bead, σ. Indeed, if a gap of size σ[3] appears between two beads, another part of the chain could pass through it, changing the topology of the chain and nullifying the effects of topological constraints. Consequently, by avoiding the creation of such a gap, we can ensure that the chain never

[1] Note that, in this framework, in this context, the excluded volume is necessary to guarantee the existence of a globule state of finite density.

[2] P.-G. de Gennes. *Scaling concepts in polymer physics*. Cornel University Press, 11 1979

[3] Actually, in 3D two gaps of size $\sqrt{2}\sigma$ may slide through each other

crosses itself. Nevertheless, the typical Rouse harmonic potential generally induces large bond fluctuations, making it difficult to guarantee the topological constraint. To avoid this, Kurt Kremer and Gary S. Grest developed a model that replaces the soft harmonic bonds, which in highly dense melts could allow for chain crossing, with very rigid links, called **finitely extensible nonlinear elastic (FENE) links**, which ensure the topological integrity of the polymer. Collapsing polymers with topological constraints will be studied in chapter 7, by means of the FENE potential.

In this chapter, however, in an attempt to isolate the different contributions to the dynamics, we want to introduce the two components necessary to make the polymer experience the coil-globule phase transition - excluded volume and attraction, **while retaining the phantom character of the chain**. To this aim, we propose a dynamical model that, as far as we know, has never been studied before, that we coined the **interacting Rouse model or iRouse model**, which adds a Lennard-Jones potential between each non-bonded monomer of a Rouse polymer, while allowing bond size fluctuations to be large enough to allow the chain to self-cross.

Our objective will then be to characterize the influence of the polymer's folding state on the iRouse polymer's dynamics. At very short times $t \ll \tau_N$ we expect to observe, exactly as for the Rouse polymer, the diffusive regime of the single monomer, as there is always a short instant during which the monomer does not yet feel the influence of its neighbors. Similarly, at long times, diffusion of the whole polymer will always dominate, taking precedence over internal dynamics. Hence, we'll mostly be interested in the intermediate regime that we will name ***internal-dynamics* regime** of the polymer, i.e. the regime in which the monomer's dynamics are influenced by the complex correlations induced by its presence within the chain. For the Rouse model, as we have seen in the previous chapter, this regime is called the Rouse regime, it is established between the relaxation times τ_N and τ_1 and it is characterized by anomalous diffusion of exponent 1/2.

In the following, our goal will be to study the **influence of volume interactions and monomer-monomer attraction** on the characteristic times and on the temporal dependence of g_1 and g_2 in the **intermediate, internal-dynamics regime for the iRouse model**.

In section 6.1, I define the iRouse model and in section 6.2, the molecular dynamics (MD) **simulation scheme** used to simulate the model. We then analyze the single monomer dynamics of the simulated polymer in the different phases - **coil** in section 6.3 and **globule** in section 6.4.

Finally, in section 6.5 we analyze the dynamics throughout the phase transition.

6.1 THE INTERACTING ROUSE - IROUSE - MODEL

6.1.1 *iRouse Hamiltonian*

The iRouse model is defined from the Rouse model by adding a Lennard-Jones potential[4] between each *non-adjacent* monomers in the chain. Let's consider an iRouse polymer of size $N+1$. Its Hamiltonian is defined as:

$$\mathcal{H}^{\text{coil}} = \mathcal{H}^{\text{Rouse}} + \sum_{i=0}^{N}\sum_{j=i}^{N} U_{i,j}^{\text{pair}} \tag{6.1}$$

where

$$\mathcal{H}^{\text{Rouse}} = \sum_{i=0}^{N-1} -\frac{1}{2}k(\vec{R}_{i+1}-\vec{R}_i)^2$$

represents the harmonic bonding interaction that ensures chain connectivity, already defined in chapter 5 for the Rouse model by equation (5.26), and

$$U_{i,j}^{\text{pair}} = \begin{cases} 4\,\epsilon_F\left(\left(\dfrac{\sigma}{r}\right)^{12} - \left(\dfrac{\sigma}{r}\right)^{6}\right) & \text{if } |j-i|>1 \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

is the Lennard-Jones pair potential that models both short-range repulsion and mid-range attraction between non-adjacent monomers.

6.1.2 *Langevin Equation for the chain*

To stay as close as possible to the Rouse model, we describe the dynamics of each bead through an overdamped Langevin. The Langevin equation for monomer $n \neq 0, N$ writes :

$$\frac{d\vec{R}_n}{dt} = -\gamma\vec{\nabla}_n[\mathcal{H}^{\text{coil}}] + \vec{\xi}_i(t) \tag{6.3}$$

where ∇_n is the three-dimensional vector operator whose three components are the corresponding partial derivative for monomer n,

$$\vec{\nabla}_i = \left(\frac{\partial}{\partial R_{i_x}}, \frac{\partial}{\partial R_{i_y}}, \frac{\partial}{\partial R_{i_z}}\right),$$

and the random forces are, as usual, delta-correlated Gaussian white noise terms obeying the fluctuation-dissipation theorem:

$$\langle \xi(t) \rangle = 0 \tag{6.4}$$

$$\langle \vec{\xi}_j(t)\cdot\vec{\xi}_i(t')\rangle = 6D\;\delta(t-t')\;\delta_{i,j}, \tag{6.5}$$

where $D = k_BT/\gamma$.

---

[4] The Lennard-Jones potential has been introduced in Figure 3.1.

6.2 LANGEVIN DYNAMICS SIMULATION SCHEME

The addition of the Lennard-Jones potential to the Rouse equation makes them non-linear, rendering the decoupling method used to solve the Rouse chain equations ineffective. There may exist approximation methods that could give some results on the dynamics of this system, but this is not the approach of this thesis. Here, as in chapter 3, I will instead analyze the result of *in silico* experiments, that I describe hereafter.

6.2.1 *Simulation Integrator: Langevin vs Brownian Dynamics*

To sample trajectories obeying the Langevin equations (6.3), we need an **integration method**, i.e an algorithmic scheme that computes the positions of all particles from one frame $\left(\vec{R}_0(t), \vec{R}_1(t), \dots \vec{R}_N(t)\right)$ to the next one $\left(\vec{R}_0(t+dt), \vec{R}_1(t+dt), \dots \quad \vec{R}_N(t+dt)\right)$, where dt is the **integration time** or **time step**.

To integrate equation (6.3) we essentially have the choice between two integration methods. A straightforward approach, known as **Brownian dynamics**, involves discretizing time in the *over*-damped Langevin equations and using Newton's method for integration: for a 1D overdamped Langevin equation, this leads to calculate, at each time step,

$$x(t+dt) = x(t) + \frac{dU(x(t))}{dx} + \xi(t).$$

Otherwise, one can integrate the *under*-damped Langevin equation (i.e. including the inertia term) using a second-order integration method, usually the velocity-Verlet algorithm[5], and then discard the unneeded short-time inertial regime. This is generally referred to as the **Langevin dynamics**.

It turns out that the Brownian dynamics are considerably less stable than the Langevin approach. Consequently, for the same simulation parameters, the Langevin integration method allows for a larger time step dt, resulting in a longer total simulation time for the same amount of computing time. For this reason, **I chose Langevin dynamics** as a simulation approach, such that I will in practice be integrating the following, underdamped, Langevin equations,

$$\begin{aligned}\frac{d\vec{R}_n}{dt} &= \vec{V}_n(t) \\ \frac{d\vec{V}_n}{dt} &= -\gamma\vec{V}_n + \vec{\nabla}_n[\mathcal{H}] + \vec{\xi}_n(t),\end{aligned} \tag{6.6}$$

even if, afterwards, our focus will be on the overdamped regime.

The selection of the **integration time** was based on empirical testing to determine the maximum value of dt that would maintain simulation stability, ultimately resulting in a value of $dt = 0.1$.

Various software exist that implement Langevin dynamics integrators. For this thesis I chose **Hoomd-Blue**[6] a molecular dynam-

---

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

[6] J. A. Anderson, J. Glaser *et al.* Hoomd-blue: A python package for high-performance molecular dynamics and hard particle monte carlo simulations. *Computational Materials Science*, 173:109363, 2020

ics package optimized for GPU computations. Hoomd integrates the Langevin equation by using the velocity-Verlet algorithm[7], which I won't explain in detail here.

6.2.2 *Simulation Physical Parameters*

All in all, our simulation depends on the **physical parameters** listed in the following table:

Parameter	**Symbol**	**Values**
The chain length	N	$[200, 400, 800, 1600, 3000, 6000, 12000]$
The Lennard-Jones dispersion energy	ϵ_F	$[0, 0.1, 0.2, 0.25, 0.27, 0.3, 0.35, 0.4, 0.45, 0.5]$
The Lennard-Jones size parameter	σ	$2^{-1/6}$
The spring stiffness	k	3
The viscosity of the fluid	γ	5
The fluid temperature parameter	k_BT	1

Table 6.1: List of the physical parameters used in MD simulations.

These parameters can be divided into two groups: the **control parameters** k, σ, γ and k_BT, which I'll leave fixed in all simulations, since their influence on the result is of no interest to us here, and the **input parameters** N **and** ϵ_F, which we'll vary to explore the different folding regimes and size effects.

As for the specific choice of the **control parameters**, our goal is to ensure that **the iRouse chain remains topologically unconstrained**, as explained in the introduction to this chapter. Consequently, the choice of the control parameters must guarantee that the bond length fluctuations are large enough to allow chain crossings. For this purpose we chose to fix $k = 3$, $k_BT = 1$, yielding $b^2 = 3k_BT/k = 1$ according to equation (1.27).

By choosing the Lennard-Jones size parameter $\sigma = 2^{-1/6} \approx 0.89$, this choice of parameters not only ensures a sufficient number of crossings to nullify reptation effects, but also sets the average link size b^2 as equal to the position of the minimum of the Lennard-Jones potential, given by $r_0 = 2^{1/6}\sigma$, giving 1 with these choices (see Figure 3.1). In this way, **the model matches qualitatively the lattice-based simulation** presented in chapter 3, in which both the equilibrium distance of the attractive interaction and the distance between two adjacent monomers in the chain were fixed by the lattice spacing, to 1. Accordingly, we expect the thermodynamic properties, and in particular the transition energy ϵ_θ, to be close to their on-lattice values determined in chapter 3.

The choice of the fluid viscosity γ is important for simulation optimization reasons. On one hand, as seen in the chapter 5, all τ_p are proportional to γ (see equation (5.41)), meaning that lowering γ accelerates the decorrelation of polymer. On the other hand, the decorrelation time of isolated monomers velocity is $\tau = \gamma^{-1}$ (equation (5.15)), i.e. the characteristic duration of the inertial regime is inversely proportional to γ, meaning lowering γ makes the inertial regime last longer. Since in our case we're not interested in the inertia regime, we want

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

to make sure it doesn't extend into the regime that characterizes the internal polymer dynamics, which runs from $t \gg \tau_N \sim \gamma$. The **optimal value I retained for the viscosity** is $\gamma = 5$, leading to $\tau = 0.2$ and $\tau_N = 0.5$.[8]

Concerning the **input parameters**, to identify potential size effects in the dynamics of the different folding states, I've chosen a wide range of polymer lengths: $N \in [200, 400, 800, 1600, 3000, 6000, 12000]$. Then, based upon the phase diagram of Figure 3.14, I chose the values of $\epsilon_F = [0., 0.1, 0.2, 0.25, 0.27, 0.3, 0.35, 0.4, 0.45, 0.5]$ that nicely sample all three phases for all values of N, as it can be deduced from the phase portrait of Figure 3.14 thanks to the expected match with the on-lattice MC simulations previously discussed.

Other parameters pertinent to the integration method can be chosen for optimization. Since these parameters don't, in principle, change the physics of the resulting simulation, I won't list them here.

6.2.3 *Simulation Integration Time Step and Total Integration Time*

In order for the simulation to cover the whole internal-dynamics regime, two requirements must be met. Firstly, the total simulation time, denoted as $T = N_{\text{steps}} dt$, should significantly exceed the longer characteristic timescale τ_1. Here, N_{steps} represents the total number of time steps. Conversely, to capture the onset of the internal-dynamics regime, the temporal resolution ΔT should be of the order of the monomer relaxation time τ_N. The temporal resolution ΔT is fixed by the acquisition period N_{period}, i.e. the number of time steps between consecutive recorded frames, given by $\Delta T = N_{\text{period}} dt$.

To meet the first requirement, for each (N, ϵ_F) I decided to run a simulation for a total $N_{\text{steps}} = 2.5\ 10^9$ time steps, which covered the interesting regime for all polymer sizes. For the second requirement, $\Delta T \approx \tau_N$, one may choose

$$\Delta T = N_{\text{period}} dt = 20 \cdot 0.1 = 2 \approx 4 \cdot \tau_N$$

meaning we should record every $N_{\text{period}} = 20$ frames.

Nevertheless, if we naively simulate $2.5\ 10^9$ frames and record every 20 frames for a $N = 12000$ polymer, the resulting files would be several terabytes in size. To manage this issue, for each (N, ϵ_F) simulation, I ran **three independent simulations, all with different total duration and temporal resolutions**, adapted to whether the focus was on the short-, medium- or long-term temporal regime. The corresponding values of N_{period} and N_{steps} are reported in the following table:

	Short time simulation	Mid time simulation	Long time simulation
N_{steps}	10^5	10^8	$2.5\ 10^9$
N_{period}	20	$2\ 10^3$	$5\ 10^4$

Table 6.2: Simulation acquisition periods and number of timesteps

In this way, the size of the records is reduced by a factor ~ 1000.

---

[8] Assuming the added Lennard-Jones pair interaction doesn't influence the value of the monomer relaxation time τ_N, which as we will see is indeed the case.

→ *Initial Particle Positions and Thermalization*

In any MD simulation, it is required to provide initial positions for each particle. Since we are interested in the equilibrium dynamics of the chain, the goal in selecting the initial position of the particles is to draw a conformation that is already as close as possible to equilibrium. In this way, as few thermalization steps as possible are required to relax the conformation and start data recording.

To achieve this, for every (N, ϵ_F) MD simulation, we conducted an on-lattice Monte Carlo simulation, as described in chapter 3, for $0.5\ N^2$ steps. This duration was empirically determined to be sufficient for equilibrating the polymer chain, as detailed in chapter 3. As mentioned above, the equilibrium state attained in the on-lattice model closely resembles that of the iRouse model. Consequently, the results obtained **from the Monte Carlo simulation were employed as the initial positions for the MD simulations**.

To ensure complete relaxation of the chain, an extra MD $50\ 10^6$ time steps were run to further equilibrate the chain.

In the following subsection, I present the simulation results and their analysis. First, for the coil and globule phase, we'll analyze the single monomer MSD by identifying the time scaling law in the internal-dynamics regime and the size scaling of the characteristic timescales. Subsequently, we will shift our attention to the evolution of the time dependence of the MSD as the system undergoes the phase transition. Our objective here is **to discern any potential critical behavior in the MSD near the phase transition point**.

6.3 COIL PHASE POLYMER DYNAMICS IN THE IROUSE MODEL

Let's start by analyzing the dynamics in the high temperature, or equivalently good quality solvent, phase where the polymer is expected to adopt the decondensed coil conformations, whose equilibrium properties were exposed in chapter 3.

Due to the low density of polymers in the coil phase, the impact of topological constraints on the dynamics of the chain should be negligible. For this reason, we can compare the dynamics of an iRouse coil polymers with existing literature about topologically constrained coil polymer dynamics. In particular the work, already mentioned in chapter 3, by D. Panja and Barkema [9], which I recall briefly.

6.3.1 *State of the art*

As mentioned in chapter 3, the dynamics of polymers in the coil phase were already investigated by D. Panja and Barkema[10].

In their paper, they conduct a Rouse mode analysis of the relaxation of on-lattice self-avoiding random walks[11]. They propose an approximate formula for the Rouse mode correlation function, based on the correlation structure of the Rouse model, i.e. an exponential

[9] D. Panja and G. T. Barkema. Rouse modes of self-avoiding flexible polymers. *The Journal of Chemical Physics*, 131(15):154903, Oct. 2009

[10] D. Panja and G. T. Barkema. Rouse modes of self-avoiding flexible polymers. *The Journal of Chemical Physics*, 131(15):154903, Oct. 2009

[11] As mentioned in chapter 3, self-avoiding random walks are a good model for polymers in the high temperature phase.

decorrelation for each Rouse mode

$$\boxed{C_{pq} = \langle \vec{X}_p(0) \cdot \vec{X}_q(t) \rangle \approx C_{pp}(0) e^{-\frac{t}{\tau_p}} \delta_{pq}}, \tag{6.7}$$

where they **postulate** the scaling for $C_{pp}(0)$

$$C_{pp}(0) = A_1 \frac{N^{2\nu}}{p^{2\nu+1}}$$

and the relaxation times

$$\tau_p = A_2 \left(\frac{N}{p} \right)^{2\nu+1}.$$

Here, $\nu = 0.588$ is the Flory exponent, already introduced in chapter 3. The constants A_1 and A_2 are introduced in their postulation **without being explicitly provided**.

From the ansatz (6.7) for the mode correlation function, the authors compute the monomer average MSD in the CM reference frame, $g_2(t)$, in the internal-dynamics regime $\tau_N \ll t \ll \tau_1$. From equation (5.54), this yields:

$$\boxed{g_2(t) \sim t^{\frac{2\nu}{2\nu+1}}.}$$

Additionally, they confirmed that the MSD of the center of mass, $g_3(t)$, exhibited normal diffusion behavior,

$$g_3(t) = 6 D_{cm} t,$$

where $D_{cm} = D/(N+1)$, in agreement with equation (5.60).

Finally, they verified that the single monomer dynamics in the lab reference frame was given by:

$$g_1(t) = g_3(t) + g_2(t)$$

as already predicted by our equation (5.52).

The validity of these relations is checked by the authors through extensive dynamic Monte Carlo simulations.

6.3.2 *Consistency Check with the iRouse Coil Polymer*

Let's confront our results with the findings by Panja and Barkema, the analysis of the Rouse mode correlation function.

→Scaling Properties of the Rouse Mode Correlation Function

Let's start by examining in detail the ansatz (6.7) made by Panja and Barkema, considering the insights provided in chapter 3 and chapter 5.

An alternative way to reach the ansatz put forward by Panja, which is not discussed in their paper, is to propose that the **Rouse modes**, despite the inclusion of excluded volume interactions, **still diagonalize the Hamiltonian** $\mathscr{H}^{\text{coil}}$ for the coil polymer, equation (6.1). If this is true, the effect of the terms added by the volume interaction would solely be to modify the scaling of the spectrum associated with the diagonalization, ϵ_p^{coil}:

$$\mathscr{H}^{\text{coil}} = \sum_{p=0}^{N} (N+1) \epsilon_p^{\text{coil}} X_p^2 \quad | \quad \epsilon_p^{\text{coil}} = A_1 \left(\frac{p}{N} \right)^{2\nu+1}. \tag{6.8}$$

This formulation, as required by the equipartition theorem, leads to the scaling of the equilibrium amplitude of the Rouse modes, which we have already encountered and explained in detail in subsection 3.3.2: it would then follow

$$\langle X_p^2 \rangle = A_1 k_B T \frac{N^{2\nu}}{p^{2\nu+1}}. \tag{6.9}$$

Furthermore, equation (6.8) implies that the Rouse modes should obey a linear stochastic differential equation of the same type as the Langevin equation (5.40), although with a different p-scaling for the friction coefficients. We therefore expect to have

$$\frac{d\vec{X}_p}{dt} = -\gamma_p \vec{X}_i + \vec{\Xi}_p(t), \tag{6.10}$$

where $\vec{\Xi}_p$ obeys the same correlation structure (equation (5.8)) as for a Rouse polymer and the friction γ_p felt by each mode is now given by:

$$\gamma_p = \frac{\epsilon_p}{\gamma} = \frac{A_1}{\gamma} \left(\frac{p}{N} \right)^{2\nu+1}. \tag{6.11}$$

Consequently, **the decorrelation times** τ_p can be expressed as:

$$\tau_p = \frac{1}{\gamma_p} = \frac{\gamma}{A_1} \left(\frac{N}{p} \right)^{2\nu+1}. \tag{6.12}$$

From equation(6.10), the correlation function for $X_p(t)$ can be computed exactly as in subsection 5.2.4, yielding the exponential correlation function postulated by Panja and Barkema, equation (6.7).

The advantage of this way of arriving at the ansatz is that it gives a new prediction on the proportionality factors A_1 and A_2 introduced by Panja.

Prediction for τ_p

Indeed, the fluctuation-dissipation theorem relative to the Rouse modes, equation (5.42), yields the important relation, which is overlooked in the original paper,

$$\langle X_p^2 \rangle = \frac{3D}{(N+1)} \tau_p, \tag{6.13}$$

or, equivalently,

$$\boxed{\frac{\langle X_p^2 \rangle}{\tau_p (N+1)} = \frac{3D}{2} = \frac{3 k_B T}{2\gamma(N+1)}}. \tag{6.14}$$

In the following section, I will verify the scaling laws (6.12), for τ_p, and equation (6.9), for $\langle X_p^2 \rangle$, and compare the prediction given by equation (6.14) to the results obtained from my simulations.

→*Numerical Results*

Following equation (6.7), I fit a decaying exponential function to $C_{pp}(t)$ obtained from the simulations, which indeed fits well $C_{pp}(t)$, as stated

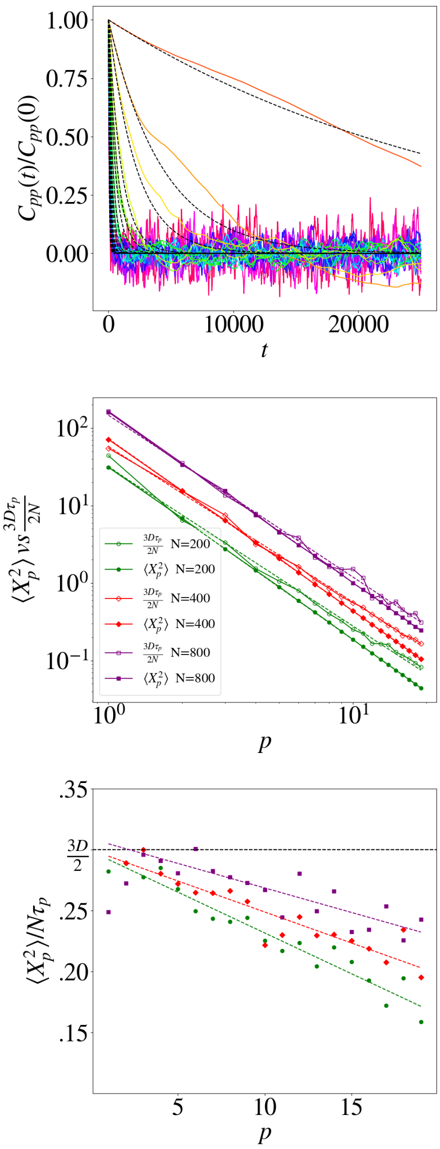


Figure 6.1: Top panel: the correlation function of mode $p \in [1 \dots 20]$ function of time, $C_{pp}(t)$. The curves are arranged in ascending order of p from right to left. The black dashed curves are exponential fits.
middle panel: $\langle X_p^2 \rangle$ and $3D/[(N+1)\tau_p]$ overlaid on the same graph. According to equation (6.14), both curves should collapse in the hypothesis that the Rouse modes diagonalize the Hamiltonian.
Bottom panel: the ratio $\frac{\langle X_p^2 \rangle}{\tau_p (N+1)}$, that should equal $3D/2 = 3/10$ given the same hypothesis.

by the authors and as shown in the top panel in Figure 6.1, so to extract the characteristic times τ_p. In the middle panel of the same figure, are overlaid $\langle X_p^2 \rangle$ and $3D/[(N+1)\tau_p]$ for the initial 20 modes of coil polymers and for three increasing chain lengths, namely $N = 200, 400, 800$, presented in log-log scale. According to equation (6.14), both quantities should collapse onto a single curve. However, by looking at the resulting plots (middle panel), we can see that, if the curves seem to match for $p \lesssim 5$, they systematically diverge from each other at larger p. Intriguingly, these discrepancies, which were mentioned in Panja et al.[58] but without precise quantification, seem to diminish with increasing polymer chain length. To quantify this behavior, I conducted power-law fits to both curves for the different values of N, yielding the following results:

$$\begin{aligned}
N = 200: &\quad \langle X_p^2 \rangle \sim p^{-2.235\pm0.007} &\qquad \tau_p \sim p^{-2.04\pm0.07}\\
N = 400: &\quad \langle X_p^2 \rangle \sim p^{-2.22\pm0.01} &\qquad \tau_p \sim p^{-2.00\pm0.04}\\
N = 800: &\quad \langle X_p^2 \rangle \sim p^{-2.20\pm0.01} &\qquad \tau_p \sim p^{-2.08\pm0.08}.
\end{aligned}$$

Interestingly, these results deviate from the expected exponent of -2.17 for both quantities. However, the exponents for $\langle X_p^2 \rangle$ and τ_p seem to be converging towards their theoretical value as N increases. This trend suggests that the **observed discrepancy may, in fact, be attributable to finite-size effects**.

Given the large error bars on the fitted exponents, to provide further support for this assertion, in the bottom panel of Figure 6.1, I plotted the ratio $\langle X_p^2 \rangle/[\tau_p(N+1)]$, which is expected to equal $3D/2$ (black horizontal line) according to equation (6.14), as a function of p. In the bottom plot of Figure 6.1, the ratio is clearly seen to converge towards the expected constant.

Coil Rouse modes correlation structure
matches findings by Panja et al.

All in all, our findings seem to **corroborate the assumptions made by Panja and Barkema**. Our added contribution is the reinterpretation of their ansatz based on the fact that the Rouse modes are still, to a good approximation, **normal modes for the coil** polymer. This important fact allows for a precise prediction of the ratio between mode amplitudes and the corresponding characteristic time scales, expressed by equation (6.14). Here we found this prediction to be correct in the large N limit, as shown in Figure 6.1, even if we highlight the presence of **finite-size effects**.

6.3.3 *Single Monomer Dynamics in the iRouse Coil*

Now, let's shift our focus to the dynamics of a single monomer. We first check the size (N) and time scaling behavior of the MSD found by Panja and Barkema by analyzing g_1 g_2 and g_3 for several values of polymer size: $N = 200, 400, 800, 1600$.

To make the visualization of our results easier to interpret, I plot the following **dimensionless quantities**,

$$\tilde{g_1} = \frac{g_1}{2R_g^2} \quad , \quad \tilde{g_2} = \frac{g_2}{2R_g^2} \quad \text{and} \quad \tilde{g_3} = \frac{g_3}{2R_g^2}, \tag{6.15}$$

as functions of the **normalized time**:

$$\tilde{t} = \frac{t}{N^{2\nu+1}} \sim \frac{t}{\tau_1}. \tag{6.16}$$

The functions $\tilde{g_1}$, $\tilde{g_2}$ and $\tilde{g_3}$ versus $\tilde{t}$ are plotted in figure Figure 6.2.

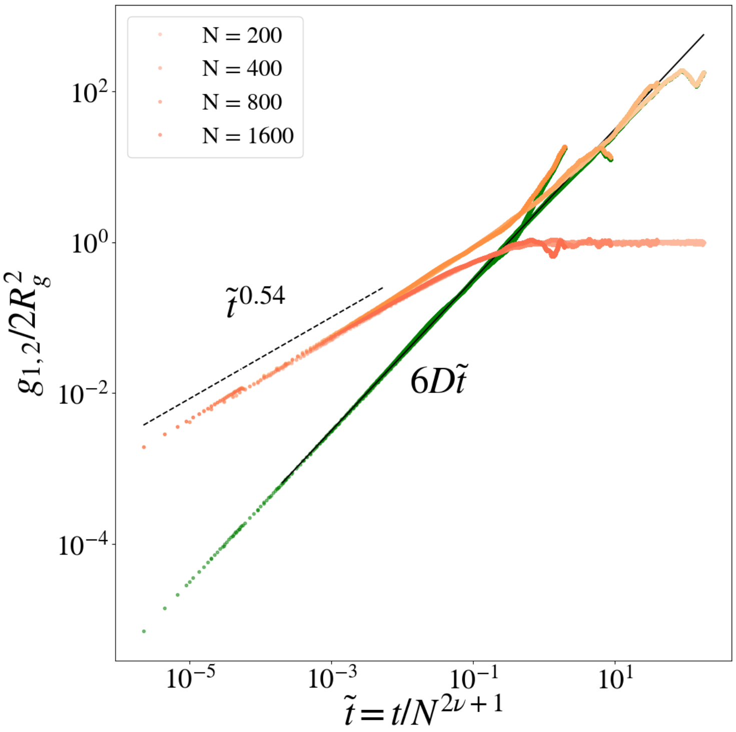


Figure 6.2: g_1 (orange graded colors) and g_2 (red graded colors) normalized by $2R_g^2$ and plotted function of $\tilde{t} = t/N^{2\nu+1}$, for $N = 200, 400, 800, 1600$. The results are similar for larger values of N, not shown here for the sake of clarity: in fact, the fluctuations are larger, as the statistics is not as good.

The observed perfect collapse of the different curves can be explained through the following key observations:

1. Since, according to equation (5.53), g_2 should plateau at $2R_g^2$ for $t \gg \tau_1$, then, in the long time regime, **all the $\tilde{g_2}$ curves must converge towards** 1:

$$\lim_{t \gg \tau_1} \tilde{g_2}(t) = 1.$$

2. Expressing $\tilde{g_1}$ and $\tilde{g_2}$ as functions of the normalized time variable $\tilde{t}$ results in **the crossovers between both time regimes occurring at $\tilde{t} = 1$.**

3. The diffusion of the CM, $g_3(t)$ for different values of N (hence the long time behavior of $g_1(t)$) perfectly collapses as well.

To explain why, first remember that, due to the free draining approximation, $g_3(t)$ should behave the same as in the Rouse case of equation (5.61), i.e. $g_3(t) = \frac{6D}{N}t$. Then, expressing $\tilde{g}_3$ as a function of the normalized time $\tilde{t}$, leads to the relation

$$\tilde{g}_3(\tilde{t}) = \frac{6D}{N} t \frac{N^{2\nu+1}}{R_g^2} = 6D\tilde{t}, \tag{6.17}$$

where **the N dependence has disappeared** due to the scaling of R_g^2 in the coil phase, given by $R_g^2 \sim N^{2\nu}$.

Taking into account the different scaling laws, Figure 6.2 finally allows for a characterization of the internal dynamics of the coil polymer:

Internal-dynamics regime of the ***coil*** *iRouse polymer*

For $t \gg \tau_N$:

1. The monomer experiences a power law **anomalous diffusion**, similarly to the Rouse polymer

$$g_2(t) \sim t^{\psi} \tag{6.18}$$

   but with **dynamic exponent** $\psi = \dfrac{2\nu}{2\nu+1} \approx 0.54$, instead of 0.5;

2. The longest relaxation time of the system, τ_1, scales as

$$\tau_1 \sim N^{\phi} \tag{6.19}$$

   with **exponent** $\phi = 2\nu + 1 \approx 2.17$, instead of 2;

3. As a consequence, **the extent of the internal-dynamics regime is increased** compared to the Rouse modes, due to the longer decorrelation time of the first mode.

4. After τ_1, the MSD in the CM reference frame, i.e $g_2(t)$, plateaus at $2R_g^2$.

These results align with and reinforce Panja and Barkema's findings regarding the single-monomer dynamics.

6.4 GLOBULE PHASE POLYMER DYNAMICS IN THE IROUSE MODEL

Let's now turn our attention to the low temperature, or equivalently poor solvent phase, where the polymer is expected to adopt the curled up **globule state**.

Let's start by recalling the equilibrium structure of the globule, reported in section 3.4. From its Rouse mode spectrum, we showed that the equilibrium globule could be seen as a **liquid of independent ideal chains**, formed by its own subchains. The linear length of subchains that can be considered as independent, s^* depends on the ratio between the chain extent and the globule size, and scales as:

$$s^* \sim N^{2/3}. \tag{6.20}$$

Figure 6.3 illustrates this point.

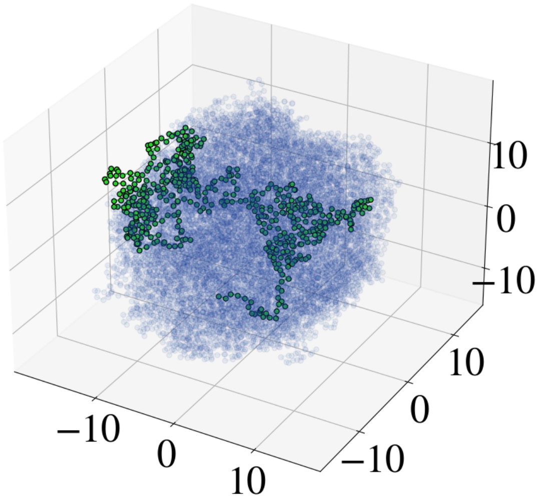


Figure 6.3: A globule conformation of length $N = 12000$ (blue) and a subchain of length $g* = 12000^{2/3} \approx 624$ (green). I claim that the relaxation process of the green subchain should obey Rouse dynamics.

What I will claim in the following is that the relaxation process of the globule *without topological constraints*, due to the Gaussian nature of the subchains, is **completely equivalent to the relaxation process of a Rouse chain of size s^*.**

My approach will be similar to that employed in the previous section. I will start with an analysis of the Rouse mode decorrelation times fitted on my simulations. This will allow for a physical understanding of the conformational dynamics in the globule. In the second part, in light of this analysis, we will investigate the single monomer dynamics.

6.4.1 *Scaling Properties of the Rouse Modes Correlation Function*

To study the scaling properties of the Rouse modes correlation function, let's *assume*, for the moment, that globules verify a number of properties previously observed for the coils. First, I'll assume that the Rouse modes diagonalize the Hamiltonian of the globule polymer (implying that their cross-correlation is zero), leading, again, to an exponential decay for the Rouse mode auto-correlation function $C_{pp}(t)$:

$$C_{pp} = \langle \vec{X}_p(0) \cdot \vec{X}_p(t) \rangle \approx \langle X_p^2 \rangle e^{-\frac{t}{\tau_p}} \,, \tag{6.21}$$

where $\langle X_p^2 \rangle$ is the spectrum of the globule, studied in detail in section 3.4, and τ_p is the relaxation time of mode p. Consequently, due to the fluctuation-dissipation theorem, the following relation is true:

$$\langle X_p^2 \rangle = \frac{3D}{2(N+1)} \tau_p. \tag{6.22}$$

These hypotheses obviously require posterior verification, as the interactions are much greater in the globule, and we may expect to observe stronger correlations among the modes.

In Figure 6.4, upper panel, I plot the correlation functions $C_{pp}(t)$. I extract values of τ_p from an exponential fit to $C_{pp}(t)$ for globule polymers for three increasing chain lengths $N = 200$, 400, 800.

Globule Rouse modes correlation structure

Interestingly, as in the case of the coil, the **exponential ansatz is found to fit well** $C_{pp}(t)$, in good agreement with our hypotheses. The resulting τ_p, along with $\langle X_p^2 \rangle$, is presented in the middle panel of Figure 6.4 for the case $N = 800$: again, τ_p **is seen to display the same** p **dependence as** $\langle X_p^2 \rangle$, which was investigated subsection 3.4.2. Hence, we observe

$$\langle X_p^2 \rangle \propto \tau_p, \tag{6.23}$$

and more specifically:

- In the high frequency modes $p > p^*$, τ_p exhibits a power law decay with exponent -2, characteristic of the Rouse polymer dynamics. Remembering that mode p represents the dynamics of sub-chains of linear length N/p, this shape for τ_p indicates that **sub-chains containing** $\boldsymbol{s^* = N/p^* \sim N^{2/3}}$ **obey Rouse like dynamics**.
- **In the first few low frequency modes** $p < p^* \sim N^{1/3}$, τ_p **is essentially constant**, meaning $\tau_1 \approx \tau_2 \cdots \approx \tau_{p^*}$. This implies that the relaxation of the whole chain, represented by τ_1, happens at approximately the same time as the relaxation of the Gaussian sub-chains at time τ_{p^*}.

All in all, this structure τ_p is in concordance with our depiction of the typical globule as a melt of independent Gaussian chains of length $s^* \sim N^{2/3}$. The relaxation process of the globule can, therefore, be broken down into the relaxation of its individual sub-chains which, independently from one another, relax through Rouse-like dynamics.

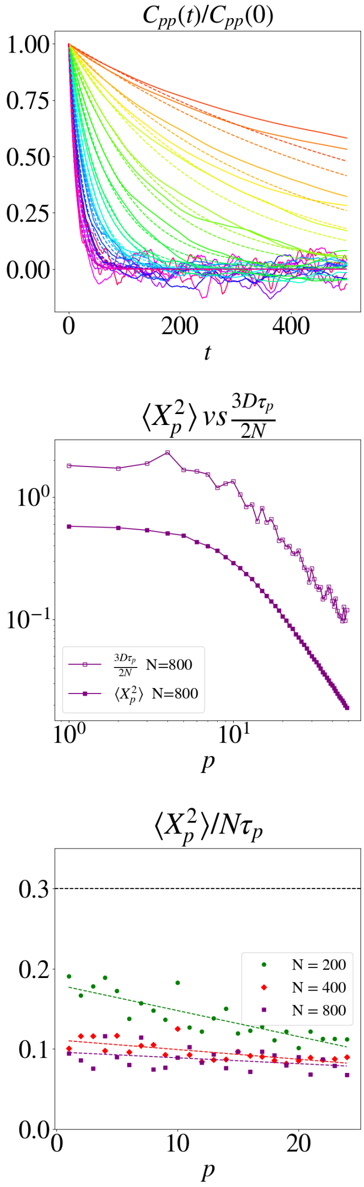


Figure 6.4: Top panel: the correlation function of mode $p \in [1 \dots 25]$ function of time, $C_{pp}(t)$ for a globule of length $N = 200$.
middle panel: $\langle X_p^2 \rangle$ and $3D/[(N+1)\tau_p]$ overlaid on the same graph for a globule of length $N = 800$. τ_p for $N = 200$ and 400 are omitted for visual clarity.
Bottom panel: the ratio $\frac{\langle X_p^2 \rangle}{\tau_p (N+1)}$, that should equal $3D/2 = 3/10$ given our hypothesis, for $N = 200$, $N = 400$, $N = 800$. As for the coil, the ratio settles to a constant function of p as N increases.

→*Proportionality Constant Discrepancy between* τ_p *and* $\langle X_p^2 \rangle$

While both $\frac{3D}{2(N+1)}\tau_p$ and $\langle X_p^2 \rangle$ are proportional, as discussed, they certainly aren't overlapped, as they should be according to our hypothesis, but only proportional, $\langle X_p^2 \rangle \propto \tau_p$. However, our initial assumptions do not accurately predict the proportionality constant. In fact, from the bottom panel of Figure 6.4 we find τ_p to be 3 times as large as our prediction, meaning that the decorrelation process of the polymer is 3 times slower than predicted by the fluctuation dissipation theorem. Given that the proportionality constant is $D/2N$, where $D = k_B T/\gamma$ is the single monomer diffusion coefficient, this is consistent with a **lower diffusion coefficient for the single monomer**.

This can be understood if we imagine a single monomer as diffusing in the globule, untethered to the chain. It is subject to two interactions, the random forces of the heat bath and the Lennard-Jones interaction with other monomers, meaning it's effectively diffusing in a Lennard-Jones fluid[12]. Hence, the lower diffusion coefficient can be seen as an effect of the crowding induced by the other monomers, modeled as a dense Lennard-Jones fluid.

6.4.2 *Single Monomer Dynamics in the iRouse Globule*

In Figure 6.5, I plot $g_1(t)$ and $g_2(t)$ for an iRouse polymer of interaction parameter $\epsilon = 0.5$ and size $N = 1600$, placing it well within the globule phase. For comparison, $g_1(t)$ and $g_2(t)$ for a coil polymer are also depicted in the same figure (in red). The black dashed line has equation $(6D/N)\ t$, which is the equation for the diffusion of the CM of both coil and globule [13].

Starting from the left of the figure, we recognize a very short inertial regime, during which the monomer doesn't feel the effect of the bonding potential and exhibits ballistic motion[14].

Following this, the internal-dynamics regime starts. For the coil, a neat power law is observed, with exponent ≈ 0.56, as stated in the previous section. In parallel, the globule also enters the **internal-dynamics regime**, also exhibiting a power law, yet with a slightly lower dynamic exponent of 0.5 - **the exponent expected for a Rouse polymer**. This is further specified by the log-log derivative of both g_1 curves, shown in Figure 6.6.

Notably, we also observe a vertical shift between both $g_2(t)$ curves in the internal-dynamics regime. This implies that the **diffusion coefficient of the monomer in the globule is lowered**, in accordance with the global shift in Rouse modes decorrelation times seen in Figure 6.4.

At larger times, the two dashed horizontal lines, at which the two g_2 functions plateaus, are at $2R_g^2$ for the coil and the globule, respectively. We immediately recognize that the plateau for the coil is much higher for the globule, simply reflecting the fact that its extent exceeds that of the globule.

Accordingly, the time at which the plateau is reached, i.e. the time at which a single monomer has diffused a distance of the order of the size of the polymer, which is none other than[15] the whole chain decorrelation time $\boldsymbol{\tau_1}$, **is much lower in the globule than in the coil**. We can estimate this time through a very simple scaling argument. For the coil

$$g_2(\tau_1) \sim {\tau_1}^{\frac{2\nu}{2\nu+1}} \sim R_g^2 \sim N^{2\nu} \tag{6.24}$$

and therefore

$$\boxed{\tau_1 \sim N^{2\nu+1}} \qquad \text{(coil)}, \tag{6.25}$$

which only confirms equation (6.19).

Now the same calculation can be made for the globule:

$$g_2(\tau_1) \sim {\tau_1}^{1/2} \sim R_g^2 \sim N^{2/3} \tag{6.26}$$

leading, this time, to

$$\boxed{\tau_1 \sim N^{4/3}} \qquad \text{(globule)}. \tag{6.27}$$

To further explore the transition from the internal-dynamics regime to the long-time regime, let's compare τ_1 to the time required for the

[12] R. Laghaei, A. Eskandari Nasrabad *et al.* Excluded volume in the generic van der waals equation of state and the self-diffusion coefficient of the lennard-jones fluid. *The Journal of chemical physics*, 124:154502, 05 2006

[13] Remember that, due to the free draining approximation, the CM always diffuses as $(6D/N)\ t$, regardless of the shape of the polymer.

[14] Note that in Figure 6.2, the inertial regime was omitted for visual clarity.

[15] Remember our discussion about the meaning of the mode relaxation times τ_p in subsection 5.2.4.

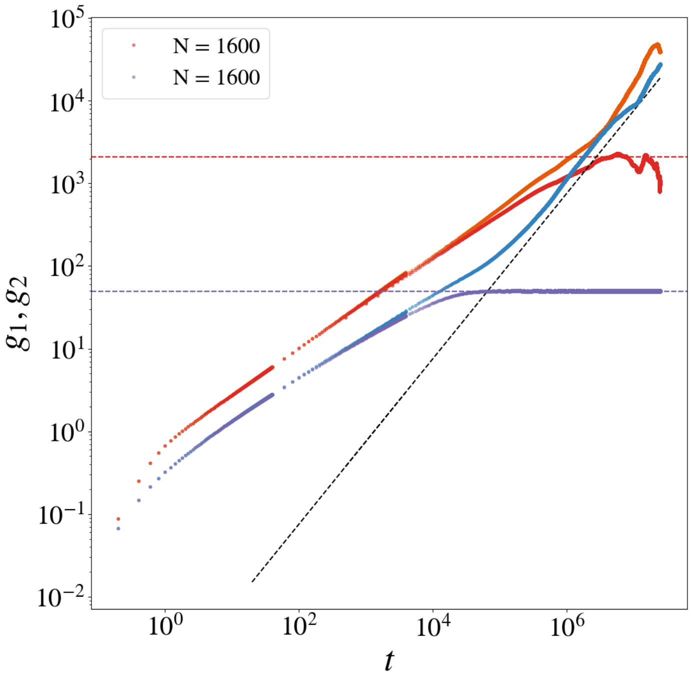


Figure 6.5: g_1 (blue) and g_2 (violet) for the globule state of a $N = 1600$ long polymer at $\epsilon = 0.5$. For comparison, in orange and red, the equivalent functions for the coil state. Dashed horizontal lines: $2R_g^2$ for the coil (red) and the globule (violet). Black dashed line: $f(t) = (6D/N)t$.

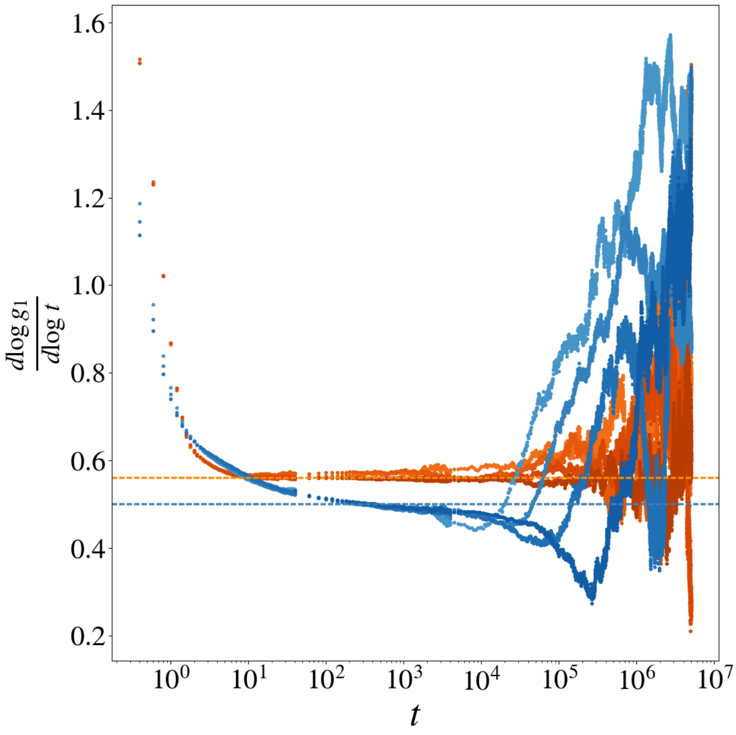


Figure 6.6: log-log derivative of $g_1(t)$ function of $\log t$. Red curves correspond to coil polymers ($\epsilon = 0$), blue curves to globule polymers ($\epsilon = 0.5$) for sizes $N = 800, 1600, 3000, 6000$. The brightness of the color indicates the size of the polymer, dark blue and red corresponding to $N = 6000$

CM itself to diffuse over a distance of the order of the chain's extent, for both the coil and the globule, starting with the coil. I denote this time by t_{cm}. Remembering that the diffusion coefficient of the CM, for both cases, scales as N^{-1}, we can write:

$$g_3(t_{cm}) \sim \frac{1}{N} t_{cm} \sim N^{2\nu} \tag{6.28}$$

yielding[16]

$$t_{cm} \sim N^{2\nu+1} \sim \tau_1 \qquad \text{(coil)}. \tag{6.29}$$

Hence, **for the coil** - and, due to the genericity of ν in this calculation, for any fractal polymer under the free draining approximation - **the time needed for a single monomer** to diffuse the extent of the whole chain is of **the same order as the time needed for the whole chain** to diffuse the **same distance**.

For the globule, that has a different size scaling, we get instead

$$g_3(t_{cm}) \sim \frac{1}{N} t_{cm} \sim N^{2/3}, \tag{6.30}$$

and thus

$$t_{cm} \sim N^{5/3} \gg N^{4/3} \sim \tau_1 \qquad \text{(globule)}. \tag{6.31}$$

In the case of the **globule, the single monomer reaches the border of the globule before** the CM diffused the same distance, i.e. before the onset of the long-time whole polymer diffusion regime. For this reason, in the lab reference frame, at $t \sim N^{4/3}$, when the monomer reaches the border of the globule, and until $t \sim N^{5/3}$, the onset of the whole polymer diffusion, the slope in g_1 is seen to dip. While quite challenging to discern in the MSD of Figure 6.5, this effect becomes evident in Figure 6.6, where the log-log derivative of the globule is presented. A dip in the log-log derivative of g_1 for the globule is clearly seen. Its intensity amplifies with N, as the difference between t_{cm} and τ_1 increases.

→*Summary of the iRouse Globule Dynamics*

To conclude this discussion, we can remark that the dynamics of a globule without topological constraints can be accurately described as a liquid of phantom Gaussian chains. The only effect of the interaction between chains is as an effective lowering of the single monomer diffusion coefficient.

From this fact I could deduce the scaling behavior of the single monomer MSD:

Internal-dynamics regime of the globule polymer: for $t \gg \tau_N$,

1. The monomer experiences a power law **anomalous diffusion**, identical to the Rouse polymer

$$g_2(t) \sim t^{\psi} \tag{6.32}$$

with **dynamic exponent** $\psi = 0.5$;

---

[16] By recalling that the Rouse mode decorrelation time τ_1 is given by equation (5.41).

2. Due the high density in the globule, the **effective viscosity** felt by a monomer **is higher** than the viscosity of the implicit solvent γ, resulting in slower dynamics.
3. The longest relaxation time of the system, τ_1, scales as

$$\tau_1 \sim N^{\phi} \tag{6.33}$$

with an **exponent** $\phi = \frac{4}{3}$, contrasting with an exponent of 2 for the Rouse polymer and $N^{2.17}$ for the coil polymer. This results in the globule's decorrelation occurring substantially faster than that observed in both the coil and the Rouse polymer. As a consequence, **the extent of the internal-dynamics regime is decreased**, due to the shorter decorrelation time of the first mode.
4. Contrary to the coil and the Rouse polymer, the onset of the long-time collective diffusion regime happens at $t_{cm} \sim N^{5/3} \gg \tau_1$. Consequently, **the beginning of a plateau is observed in** g_1, resulting in a dip in its log-log derivative.

6.5 DYNAMICS THROUGHOUT THE TRANSITION AND CONCLUDING REMARKS ABOUT THE SPECTRAL MODES APPROACH

In addition to characterizing the dynamics of the interacting polymers, our aim was also to understand whether the dynamics through the different phases of the model give the possibility of discerning the folding state of the polymer based on the behavior of an individual monomer. Now that I have characterized the dynamics in both the coil and the globule phase, I will therefore, in this section, investigate more explicitly how the observed dynamical features evolve across the coil-globule phase transition.

→MSD Results

To get a more precise idea of the dynamics along the transition, I ran simulations for iRouse polymers for $N = 400$ and $\epsilon = [0, 0.1, 0.2, 0.25, 0.27, 0.3, 0.35, 0.4, 0.45, 0.5]$. The MSD in the center of mass reference frame, $g_2(t)$, for a polymer of fixed length at different interaction energy parameters ϵ crossing the transition point are plotted in Figure 6.7. The color of each curve is based on the order parameter α for the phase transition defined in chapter 3, so that α spans from -2.2 for coils to 0 for globules. A few comments on this figure help to summarize and clarify the previous results.

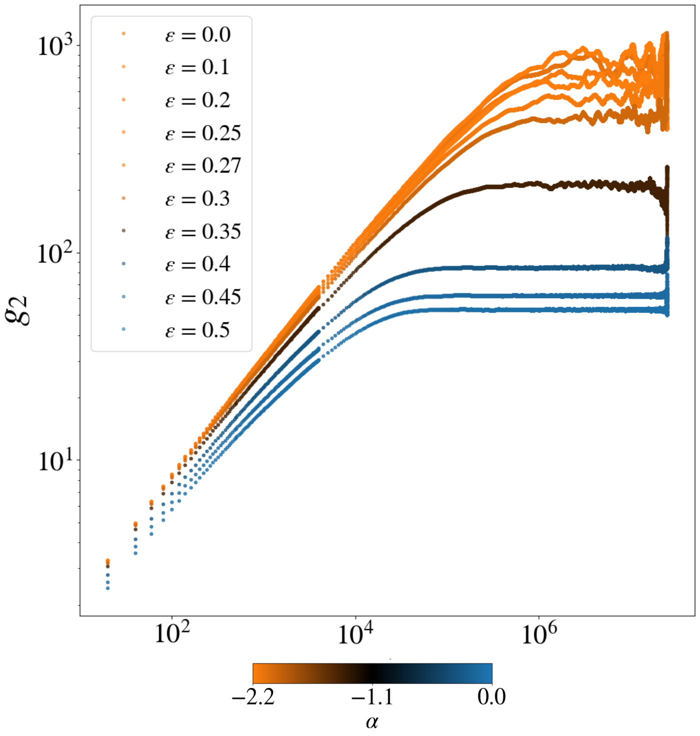


Figure 6.7: MSD in the center of mass reference frame, for polymers of length $N = 400$ and varying ϵ.

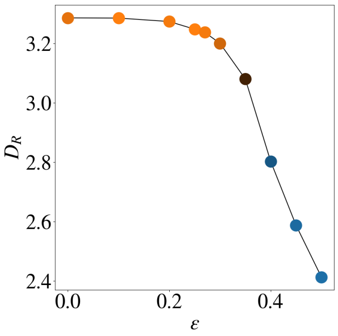


Figure 6.8: The diffusion coefficient D_{R} fitted on the corresponding $g_2(t)$ of Figure 6.7 by mean of a $D_{\mathrm{R}} t^{\psi}$ power law.

MSD scaling throughout the coil-globule phase transition

- First of all, the lowering of the long time **asymptotic value** of $g_2(t)$ as ϵ increases is a testimony to the collapse of the polymer (as it correspond to $2R_g^2$).
- Accordingly, **the typical time** τ_1 needed for the single monomer to reach this limit, i.e. to visit the whole polymer, also decreases, as shown by the shift of the curve's shoulder toward the left. This aspect will be explored hereafter.
- As the collapse occurs, the density in the polymer increases, leading to **the diffusion coefficient** D_{R} in the Rouse regime to decrease as well. This phenomenon is depicted in Figure 6.8, showing the D_{R} parameter fitted on the corresponding $g_2(t) \sim D_{\mathrm{R}} t^{\psi}$. As discussed, this effect is equivalent to an increase in the effective viscosity.
- Concurrently, **the dynamic exponent** ψ of the MSD in the internal dynamics regime decreases continuously from its coil value of ≈ 0.56 to its globule value of 0.5, as depicted in Figure 6.9:

$$\boxed{\psi^{\text{coil}} = 0.56 \qquad | \qquad \psi^{\text{globule}} = 0.5.} \tag{6.34}$$

Throughout the transition, ψ evolves continuously from one value to the other.

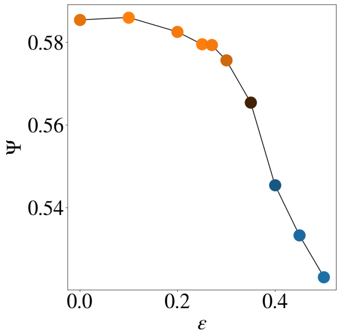


Figure 6.9: The power law exponent ψ fitted on the corresponding $g_2(t)$ of Figure 6.7 by mean of a $D_{\mathrm{R}} t^{\psi}$ power law.

go to ToC

→*Spectral Interpretation and Fluctuation Dissipation Theorem*

To better interpret the shape of the MSDs, I turned again to the modes, and more specifically to their decorrelation time.

As before, I fitted an exponential to the correlation functions $C_{pp}(t)$ to extract the decorrelation times τ_p. The results are shown on Figure 6.10, side-by-side with the spectra $\langle X_p^2 \rangle$ for the same simulations.

In the previous sections, we have established that, as a consequence of the **fluctuation dissipation theorem, τ_p and $\langle X_p^2 \rangle$ were proportional**. From Figure 6.10, we observe that this proportionality remains **throughout all stages of the coil-globule phase transition**. Consequently, the same analysis detailing the evolution of $\langle X_p^2 \rangle$ in subsection 3.5.1 can be applied to describe the evolution of τ_p. In particular, we notice in Figure 6.10 that the value of the longest relaxation time of the polymer, τ_1, is seen to **abruptly decrease** as the transition takes place, and the first few relaxation times converge **toward a same value**.

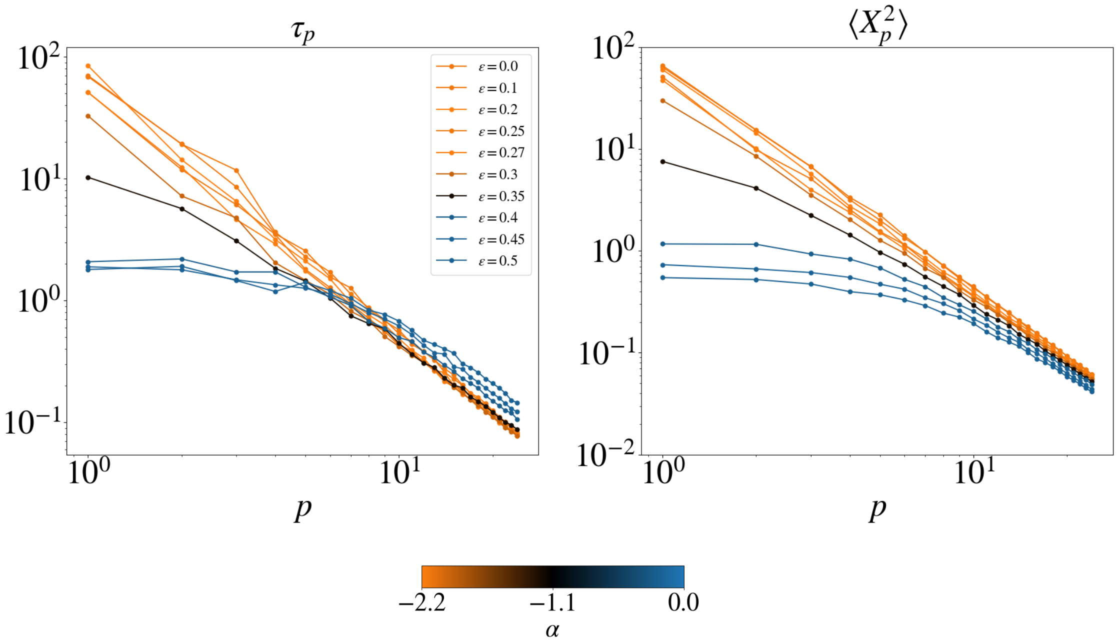


Figure 6.10: Left panel: the evolution of the decorrelation times τ_p of polymers at fixed length $N = 400$ for epsilon values across the coil globule phase transition.
Right panel: Rouse modes spectra for the same polymers.

Also, as the polymer enters the globule phase, the curves for τ_p experience an upward shift, indicating an evolution of the proportionality constant between both observable τ_p and $\langle X_p^2 \rangle$, mirroring the decrease of D_{R} previously observed. As stated in the analysis of the globule dynamics, this shift is due to the **increase in effective viscosity** felt by the monomers getting greater due to the increase in monomer density.

6.6 CONCLUSION/TAKE-HOME MESSAGE

In this chapter, I presented a simulation study of an original model, the **interacting Rouse model**, that incorporates all equilibrium aspects of the coil-globule phase transition. The particularity of this model relies in the fact that it is allowed to self-cross, which influences its dynamics. I characterized the dynamics of the model in all folding states by performing an analysis of both the relaxation processes of the chain and the single monomer dynamics. Let's summarize our findings.

Our key result in this chapter is that the **proportionality relation** between the decorrelation time τ_p and its associated Rouse mode amplitude $\langle X_p^2 \rangle$, encountered in the case of the Rouse polymer, remains true for polymers at any folding states, thereby establishing a direct connection between the chain's spatial structure and its decorrelation process. This relation stems from the fact that, regardless of volume interactions, the Rouse modes seem to **diagonalize the equations of motion of the system**, exactly as they did in the case of the Rouse model, allowing for the fluctuation-dissipation to be applied to each mode, yielding the proportionality relation.

This relation helped us to elucidate the **dynamics of a single monomer** in all folding states - coil, globule, and transition - which we found to **mirror the structure observed in a Rouse polymer**. Specifically, we observe an internal dynamics regime where the monomer undergoes subdiffusion because of the chain's influence. This lasts until a characteristic time, τ_1, which is the polymer's longest relaxation time. After this time, the monomer displays regular diffusion, moving coherently with the rest of the chain. Strikingly, therefore, the decorrelation times τ_p appear as the dynamic equivalent of the spectral modes, as nicely illustrated by Figure 6.10, with lower modes representing an excellent observable in assessing the polymer state.

The **dynamic exponent** ψ, characterizing the subdiffusion in the Rouse regime, changes only **slightly** from one phase to the other, going from its coil value of ≈ 0.56 to its globule value of 0.5, and hardly departs from its Rouse value of 0.5. Between both phases, ψ evolves continuously from one value to the other. This dynamic exponent can be seen as a consequence of the scaling of the decorrelation spectrum τ_p, as attested by the equation (5.54), linking the correlation function to the MSD.

On the contrary, **the extent of the Rouse regime**, which is controlled by τ_1, **changes dramatically** as the polymer collapses. It goes from a scaling of $\approx N^{2.17}$ in the coil phase to a scaling of $\approx N^{1.33}$ in the globule phase. This can be seen simply as a consequence of the size of the polymer reducing during the collapse.

Finally, we found that, due to the high monomer density in the globule phase, the **effective viscosity** felt by the monomer increases.

All in all, the crucial message to take home from this chapter is that, despite the intensity of the volume interactions, if topological constraints are neglected, **the internal dynamics of the chain re-**

main essentially *Rouse-like*, with only a slight change in the dynamical exponent. Notably, **nothing extraordinary happens at the phase transition**. Indeed, one could have expected that the divergence of the order parameter fluctuations described in chapter 3 and displayed in Figure 3.13 and the animation Figure 4.7, could potentially translate into some salient feature appearing in the dynamics of the chain, but this is not the case.

In the following chapter, we impose topological constraints on the chain and analyze how this affects the chain dynamics.

7

Dynamics of a collapsing polymer with topological constraints

7.1 INTRODUCTION

In the previous chapter, I introduced a polymer physics model designed to study the dynamics of the coil-globule phase transition without topological constraints, thereby essentially exhibiting the dynamics observed in the Rouse model. This offered a streamlined perspective, allowing for a focused exploration of the dynamics inherent to the phase transition without the additional complexity introduced by **polymer entanglement**.

As we progress into this chapter, I expand the scope of investigation to reintegrate topological constraints into the model. Notably, the incorporation of these constraints introduces the expectation of observing **reptation dynamics**, a theory that postulates that a polymer, when constrained by its entanglement with other polymer chains, will navigate through a 'tube' defined by the chain's entanglements, via a snake-like movement, thereby greatly slowing down its relaxation dynamics.

The objective will be to examine the **polymer dynamics** across all phases of the coil-globule transition in the **presence of topological constraints** and to compare these findings with the dynamics of the unconstrained model. We can reasonably expect that in the coil phase, characterized by rare chain self-collisions, these constraints might not play a substantial role. In contrast, within the dense globule phase, the importance of such constraints is paramount, and has already been observed[1]. Hence, I'll aim in this chapter, by comparison to the unconstrained case, to assess the physical parameters, i.e. the (N, ϵ) region, where topological constraints do impact the dynamics.

Although a thorough theoretical and simulation exploration of the dynamics in the globule phase, based on reptation theory, is a fascinating subject in itself, to my surprise it has never been undertaken. I won't carry it out completely in this thesis, but I will begin to address it, by limiting myself to the study of qualitative differences between the

[1] M. Tamm, L. Nazarov *et al.* Anomalous diffusion in fractal globules. *Physical review letters*, 114, 04 2014

topologically constrained and unconstrained cases. This will at least enable me to assess when reptation is or is not present across the (N, ϵ) coil-globule phase diagram, which may, however, have interesting consequences for the analysis of experimental data.

The outline of the chapter is as follows. In section 7.2, I begin by briefly presenting my **simulation scheme** for simulating topologically constrained polymers, based on the Kremer-Grest model with added attractive interaction. Then, in section 7.3, I briefly review the **theory** of dynamics in topologically constrained polymer solutions.

Eventually, in section 7.4 and section 7.5, by comparing the dynamics in the constrained and unconstrained case, I assess the (N, ϵ) **parameters required for topological constraints to impact** the dynamics.

7.2 SIMULATION OF A TOPOLOGICALLY CONSTRAINED INTERACTING POLYMER: THE KREMER-GREST MODEL WITH ATTRACTIVE LENNARD-JONES INTERACTIONS

7.2.1 *The Kremer-Grest model*

In the previous chapter, we employed a harmonic bonding potential to simulate chain connectivity. This allowed for the chain to perform self-crossings, prohibiting the onset of reptation dynamics. The reason for this is that the harmonic potential is "soft", in the sense that it allows for large size fluctuations. To avoid this, Kurt Kremer and Gary S. Grest established a new model[2] for the bonding interaction, named **finitely extensible nonlinear elastic** (FENE) potential, that reads:

$$U_{\mathrm{FENE}}(r) = -\frac{1}{2} k R_0^2 \ln\left(1 - \left(\frac{r}{R_0}\right)\right) + W_{\mathrm{WCA}}. \tag{7.1}$$

Here, k is the bond strength, and R_0, the maximum extendable length of the bond, is ensured by the **exponential divergence of the potential in $r = R_0$**. The additional U_{WCA} term is the Weeks-Chandler-Andersen (WCA) potential, a shifted and truncated form of the Lennard-Jones (LJ) potential, retaining only its repulsive part. It reads:

$$U_{\mathrm{WCA}}(r) = \begin{cases} U_{\mathrm{LJ}}(r) - U_{\mathrm{LJ}}(r_c) & \text{if } r < r_c \\ 0 & \text{otherwise} \end{cases} \tag{7.2}$$

where the LJ potential is defined as usual, by

$$U_{\mathrm{LJ}}(r) = 4\epsilon_{\mathrm{WCA}} \left[\left(\frac{\sigma_{\mathrm{WCA}}}{r}\right)^{12} - \left(\frac{\sigma_{\mathrm{WCA}}}{r}\right)^{6} \right].$$

Crucially, the non-crossing condition depends on the selection of the potential parameters, which the Kremer-Grest model suggests choosing as follows:

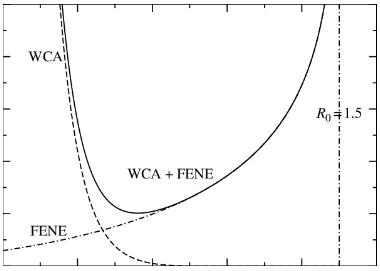


Figure 7.1: FENE and WCA potentials. Adapted from Ref. [37].

---

[2] K. Kremer and G. S. Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086, 04 1990

R_0	k	ϵ_{WCA}
$1.5\ \sigma_{\text{WCA}}$	30	1

Overall, the use of this bonding potential has consistently demonstrated its efficacy in simulating polymers for which **chain crossing is prevented**. To complete our model, we only need to incorporate an attractive pair interaction between all non-bonded monomers.

7.2.2 *The Interacting Kremer-Grest Model*

My simulation scheme is then exactly the same as for the iRouse model simulations of the preceding chapter, except that the harmonic bonds are replaced by **FENE bonds**. The Hamiltonian of the system hence reads:

$$\mathcal{H}^{\text{coil}} = \sum_{i=0}^{N-1} U_{\text{FENE}}(r_{i,i+1}) + \sum_{i=0}^{N} \sum_{j=i}^{N} U_{i,j}^{\text{pair}} \tag{7.3}$$

where $U_{i,j}^{\text{pair}}$ is the Lennard-Jones pair potential of equation (6.2) that again models short-range repulsion and mid-range attraction between *non-adjacent* monomers.

As before, the system is in contact with a heat bath at temperature $k_B T = 1$, modeled by a stochastic force following the Langevin dynamics formalism. The equation of motion for each bead is

$$\frac{d\vec{R}_n}{dt} = \vec{V}_n(t)$$
$$\frac{d\vec{V}_n}{dt} = -\gamma \vec{V}_n + \vec{\nabla}_n[\mathcal{H}] + \vec{\xi}_n(t),$$

exactly as for the iRouse model, see equation (6.6).

I integrated these equations using the Langevin integrator of Hoomd-Blue. All relevant simulation parameters are listed in the following table:

Parameter	**Symbol**	**Values**
The chain length	N	$[200, 400, 800, 1600, 3000, 6000, 12000]$
The Lennard-Jones dispersion energy	ϵ_F	$[0, 0.1, 0.2, 0.25, 0.27, 0.3, 0.35, 0.4, 0.45, 0.5]$
The Lennard-Jones size parameter	σ	$2^{-1/6}$
The WCA size parameter	σ_{WCA}	$2^{-1/6}$
The WCA dispersion energy	ϵ_{WCA}	1
The FENE bond strength	k	30
The FENE maximum length	R_0	$1.5\sigma_{\text{WCA}}$
The viscosity of the fluid	γ	5
The fluid temperature parameter	$k_B T$	1

Table 7.1: List of the physical parameters used in the interacting Kremer-Grest model MD simulations.

7.3 Polymer Motion in an Entangled Polymer Solution: Reptation Dynamics

Although the behavior of polymers in dilute solutions, where entanglements are absent, is well understood[3], the complex dynamics introduced by entanglement complexities remain a topic of ongoing debate and exploration. As mentioned in the introduction, the aim of this chapter isn't to conduct a detailed characterization of the effect of entanglement itself on the polymer dynamics. While such an endeavor would undoubtedly be of great academic interest, our focus here is more modest and is to specify the limiting conditions for the onset of entanglement effects. In this spirit, I won't explain in detail the complicated physics underlying the subject, but rather give a qualitative picture of the single monomer dynamics in the presence of entanglement effect, and the conditions making them appear.

→*The reptation tube*

In brief, then, the idea behind the reptation motion lies in the definition of an idealized constraining **tube**, enveloping each polymer chain in the solution, that models the steric interaction with its surroundings. For substantial movement to occur, the chain must navigate longitudinally out of this tube.

Meanwhile, the configuration of the tube is susceptible to change via two mechanisms. Initially, as the central chain moves, it abandons sections of its initial tube while creating new segments. Concurrently, the tube's form can also fluctuate due to the movements of surrounding chains that constitute the tube. A typical **tube lifetime** can therefore be defined, τ_{tube}. The key point is that, **for sufficiently long chains**, in a certain time regime that we shall specify shortly, the tube fluctuations are negligible, meaning the chain motion is dominated by its longitudinal motion, coined **reptation motion**.

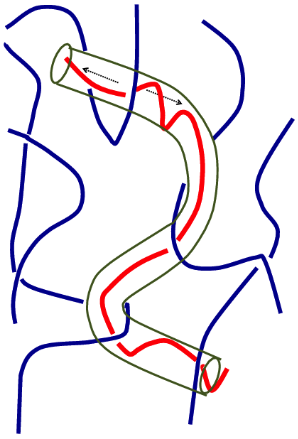


Figure 7.2: Schematic representation of the reptation tube.

→*Minimal Polymer Length for Entanglement*

Determining the minimal polymer length at which entanglement effects are important is a very complex and still debated question. Several theoretical formulas have been proposed for different concentration conditions [4,5]. The most important point is that, quite intuitively, **the minimum length N_e is a decreasing function of monomer density** $\rho = N/R_g^3$ in the solvent. This function is as a power law, with both the proportionality constant and the exponent being subjects of ongoing debate.

The determination of N_e is a fascinating subject, but as we said earlier, we aren't going to dwell on it. For our discussion, we only require an understanding of the qualitative relationship between N_e and monomer density: as the solution becomes denser, shorter chains are necessary for the entanglement effect to manifest.

[3] M. Doi and S. Edwards. *The Theory of Polymer Dynamics*. International series of monographs on physics. Clarendon Press, 1988

[4] S. T. Milner. Unified entanglement scaling for flexible, semiflexible, and stiff polymer melts and solutions. *Macromolecules*, 53(4):1314–1325, 2020

[5] T. A. Kavassalis and J. Noolandi. New view of entanglements in dense polymer systems. *Phys. Rev. Lett.*, 59:2674–2677, Dec 1987

→*Chain Dynamics in Entangled Solutions*

In the case of solutions where the polymer chain length N is less than the entanglement length N_e - meaning the dilute solution limit or, equivalently, short chain limit - the lifetime τ_{tube} of the effective tube is much shorter than the relaxation time of the chains, τ_1. Consequently, the dynamics of the chain are described by the dynamics of the equivalent isolated chain, Rouse dynamics for short concentrated polymer solutions, coil dynamics for dilute long chains.

Now let's consider a solution with $N > N_e$. In this case, $\tau_{\text{tube}} \gg \tau_1$, meaning that for times $t \ll \tau_{\text{tube}}$, the tube can be seen as a fixed constraint for the polymer. This constraint introduces a new length-scale, the **tube diameter** d, and a **corresponding time-scale** τ_e, the time needed for a monomer to move a length d. The motion of the chain, and thus of a single monomer, can then be decomposed into **five time regimes**[6]:

1. For $t \ll \tau_N$ The monomer doesn't feel the presence of the chain and displays normal diffusion.
2. For $t \ll \tau_e$: The monomer feels the presence of the chain, but doesn't feel the presence of its surrounding and the chain experiences **Rouse dynamics** within the tube diameter, appointing the Rouse exponent for the single monomer MSD:
$$g_1(t) \sim t^{1/2}. \tag{7.4}$$
3. For $\tau_e \ll t \ll \tau_1$: The monomer is constrained to move up and down the tube through reptation motion. The dynamics of the chain in this regime can be seen as a **Rouse-like diffusion along the path of the tube**, since $t \ll \tau_1$. As the tube is expected to have Gaussian conformations $s^{1/2}$, the single monomer scaling is
$$\boxed{g_1(t) \sim \left(t^{1/2}\right)^{1/2} \sim t^{1/4}}, \tag{7.5}$$
predicting a slowing down of the dynamics due to the topological constraints. This is the so-called **reptation regime**.
4. For $\tau_1 \ll t \ll \tau_{\text{tube}}$: The bead still moves along the tube. Now, however, $t > \tau_1$, which means the monomer moves coherently with the whole chain which experiences **normal diffusion along the path of the tube**, yielding
$$g_1 \sim t^{1/2}. \tag{7.6}$$
5. Finally, for $t > \tau_{\text{tube}}$: the tube itself starts diffusing, along with the polymer, meaning
$$g_1(t) \sim t. \tag{7.7}$$

Importantly, for the reptation regime to be observed, very long simulations of very long polymers in concentrated conditions (i.e. by simulating a large number of polymers in a given volume) must be performed.

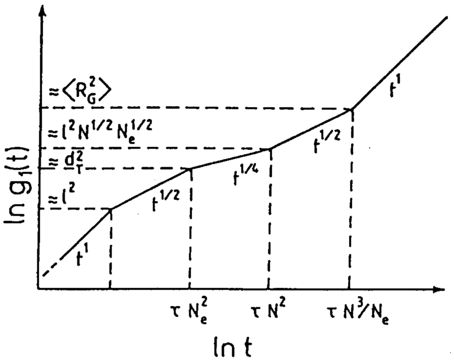


Figure 7.3: Theoretical $g_1(t)$ as predicted by tube theory from Ref. [44].

---

[6] M. Doi and S. Edwards. *The Theory of Polymer Dynamics*. International series of monographs on physics. Clarendon Press, 1988

[7] H.-P. Hsu and K. Kremer. Static and dynamic properties of large polymer melts in equilibrium. *The Journal of Chemical Physics*, 144(15):154907, 04 2016

This has been realized in particular by Hsu and Kremer in Ref. [7]. The resulting plot of $g_1(t)$, for a solution of $M = 1000$ chains of length $N = 500$ and 2000 is presented in Figure 7.4. While, in this figure, all but the whole tube diffusion is visible, as the length of the polymer decreases, so does the tube lifetime τ_{tube}, shortening the duration of regimes 3 and eventually the reptation regime. This continues until $N < N_e$ at which point only the Rouse regime is left, and dilute solution dynamics are retrieved.

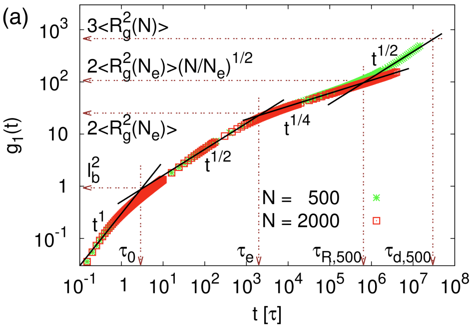


Figure 7.4: $g_1(t)$ for dense polymer melt of chains of size $N = 500$ (green) and $N = 2000$ (red) fom Ref. [36]

→Reptation Dynamics in Topologically Constrained Polymer Across the Coil-Globule Phase Transition

As mentioned in the previous subsection, the impacts of entanglements and the consequent emergence of reptation dynamics are usually explored in the realm of polymer solutions, where the entanglements of a chain are mostly due to the other chains of the solution. However, in the case that interests us, of a single polymer, the entanglements will be with other parts of the same chain. Hence, the chain length to be compared with the entanglement length **isn't the length of the full polymer** N, but some length $s^* < N$, such that subchains of length s^* can be considered as independent. As we learned from the spectrum of the globule conformations, this corresponds to the correlation length of the chain. Consequently, the full polymer can be modeled **as a solution of $M = N/s^*$ independent subchains**.

Note that we have a prediction for s^* in the case of a perfect globule, for which the decorrelation length goes as $s^* \sim N^{2/3}$, implying $M \sim N^{1/3}$. Otherwise, especially in the transition and in the soft globule phase, we expect s^* to be a nontrivial function of both the full polymer size N and the degree of compaction of the chain, controlled by the attractive interaction parameter ϵ. Then, at fixed ϵ, the degree of compaction itself is dependent on the size of the polymer, due to the collective nature of the transition (see Figure 3.14). Finally, the entanglement length N_e in turn depends on the compaction. All in all, the intensity of the entanglement effect in the single polymer depends on the complex interplay between chain length, chain compaction, and chain correlation length. Consequently, it seems quite challenging to produce a scaling theory for reptation dynamics in the single polymer.

Nevertheless, in the following section, I present the first steps towards this objective, by identifying empirically in numerical simulations, the threshold values of N and ϵ above which topological constraints can have an effect on the dynamics.

7.4 ASSESSING THE IMPACT OF TOPOLOGICAL CONSTRAINTS ON INTERACTING POLYMER DYNAMICS

7.4.1 *Equilibrium Properties: Phase Diagram and Size Difference Between Both Models*

Remember that **equilibrium scaling** properties shouldn't be dependent on whether the chain is topologically constrained or not. Nevertheless, the modification of the bonding potential does influence the value of certain equilibrium features of the chain.

Notably, the equilibrium size of the chain for both models, characterized by the gyration radius R_g, is plotted in Figure 7.5.

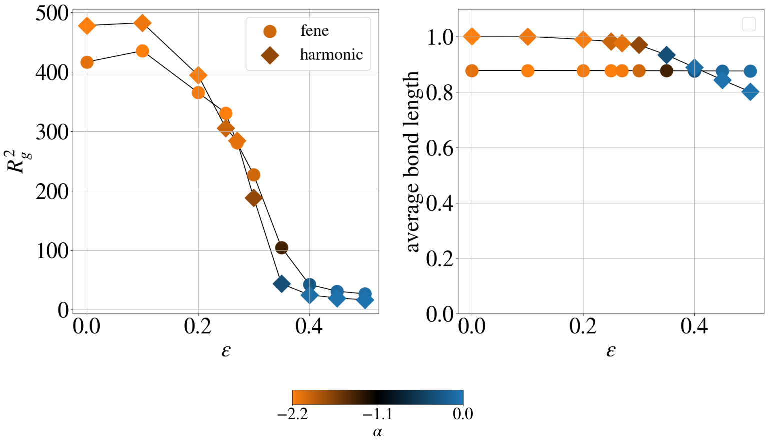


Figure 7.5: Left panel: The gyration radius, at fixed $N = 200$ and different values of ϵ across the coil-globule phase transition, for both models. Diamonds and dots represent the iRouse and FENE polymers, respectively. Again, the color of the points is based on the order parameter α, orange for the coil phase, blue for the globule phase, and black for the transition.
Right panel: The average bond length for the same polymers.

→*Size difference*

The first observation that we can make is that, in the coil phase, the FENE chain is larger than the harmonic chain. This is simply due to the fact that, given the choice of parameters, the mean bond length in the coil phase is longer for the harmonic polymer compared to the FENE polymer. The opposite is true in the globule phase: both effects are visible on the plot of the average bond length as a function of ϵ, shown in the second panel of Figure 7.5. While the average FENE bond length stays rigorously constant, in the case of the harmonic bond length, it is seen to decrease as the polymer collapses. This is due to the fact that no excluded volume interaction is introduced between harmonically bonded monomers in the iRouse model, allowing for neighboring monomers to slightly overlap due to the rising pressure as the density increases in the globule phase.

→*Phase diagrams*

In Figure 7.6, I again make use of the order parameter α introduced in chapter 3, to draw the phase diagram for both models. The color of the dot at position (N, ϵ) represents the value of α for this pair of parameters, exactly as in Figure 7.6. Orange, Blue, and Black respectively represent $\alpha = -2.2, -1.1, 0$, i.e. the coil, globule, and transition phase.

However, the resolution (N, ϵ) of these diagrams is significantly lower than for the iSAW model Figure 3.14, due to the much higher computational cost of Langevin simulations compared to the Monte Carlo simulation performed for the on-lattice model in chapter 3, limiting the number of simulated (N, ϵ) points. Also noteworthy, the uncertainty on the value of α increases with N, and is worse in the coil phase than in the globule phase. This is a consequence of the polymer's decorrelation time scaling as $\approx N^{2.17}$ in the coil phase and $\approx N^{4/3}$ in the globule phase, whereas the number of simulation steps remains constant for each simulation.

Nevertheless, a qualitative picture of the phase diagram remains discernible. For comparison, the inflection point critical line, defined in (3.17), is drawn in the background and is seen to align with the transition region for both models. Hence the key result is that all three models **share essentially the same phase diagram**.

As we'll see, these observations will prove useful to understand the behavior of the single monomer dynamics in the next section.

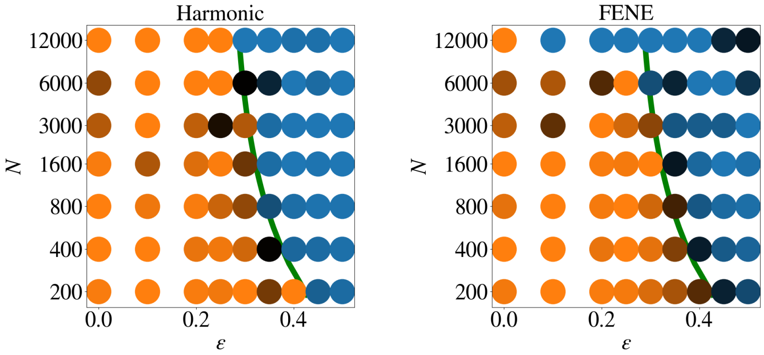


Figure 7.6: Finite-size coil-globule phase diagrams for the Harmonic (left) and FENE (right) drawn from a color plot of the order parameter α. The green line in the background represents the inflection point (defined in subsection 3.5.2).

7.4.2 *Simulation Results for the Monomer MSD*

We can now compare the single monomer dynamics in a polymer with and without topological constraints. As a point of comparison, I choose the observable g_2, which isolates best the internal dynamics regime.

In Figure 7.7 (top two rows) is depicted for $N = 3000$ and varying ϵ values across all stages of the collapse. The colored curves correspond to the FENE bond simulations, while the thicker gray curves correspond to the harmonic bond simulations.

In the coil phase, $\epsilon \leq 0.3$, the $g_2(t)$ curves are **essentially identical** for both models, meaning topological constraints have no effects on the dynamics of the coil polymer, as expected, due to the low density of coil[8].

Around $\epsilon = 0.35$ the collapse occurs; hence, from $\epsilon \geq 0.35$, the R_g difference between both models is substantial, as pointed out in the previous section. Consequently, the time needed for the monomer to reach the plateau at $2R_g^2$, i.e. the chain decorrelation time τ_1, is lower in the harmonic bond simulation. Nevertheless, the short-time dynamics, which correspond to the Rouse regime, are seen to match nicely, apart from a slightly higher decay of the diffusion coefficient in globule without topological constraint ($\epsilon = 0.5$).

To facilitate a visual comparison, I chose to normalize the $g_2(t)$ curves in order to align the plateau values, by dividing $g_2(t)$ by the fitted $2R_g^2$,

$$\tilde{g}_2 = \frac{g_2}{2R_g^2}. \tag{7.8}$$

This, of course, will misalign the Rouse dynamics regime, hence I also rescale the time axis, in order to force the Rouse regimes to collapse for both simulations:

$$t' = t\left(\frac{D_R}{2R_g^2}\right)^{1/\Psi} \tag{7.9}$$

where D_R and Ψ are the exponent and prefactor of g_2 fitted in the Rouse regime, respectively. The curves for $\tilde{g}_2(t')$, are represented in Figure 7.7 (bottom two row).

Thanks to this precise realignment of the curves, in this figure a lowering of the dynamic exponent for intermediate times is clearly visible from $\epsilon = 0.35$ and beyond, indicating the **onset of reptation dynamics** as the density in the globule rises.

[8] Discrepancies at long times are due to poor statistics.

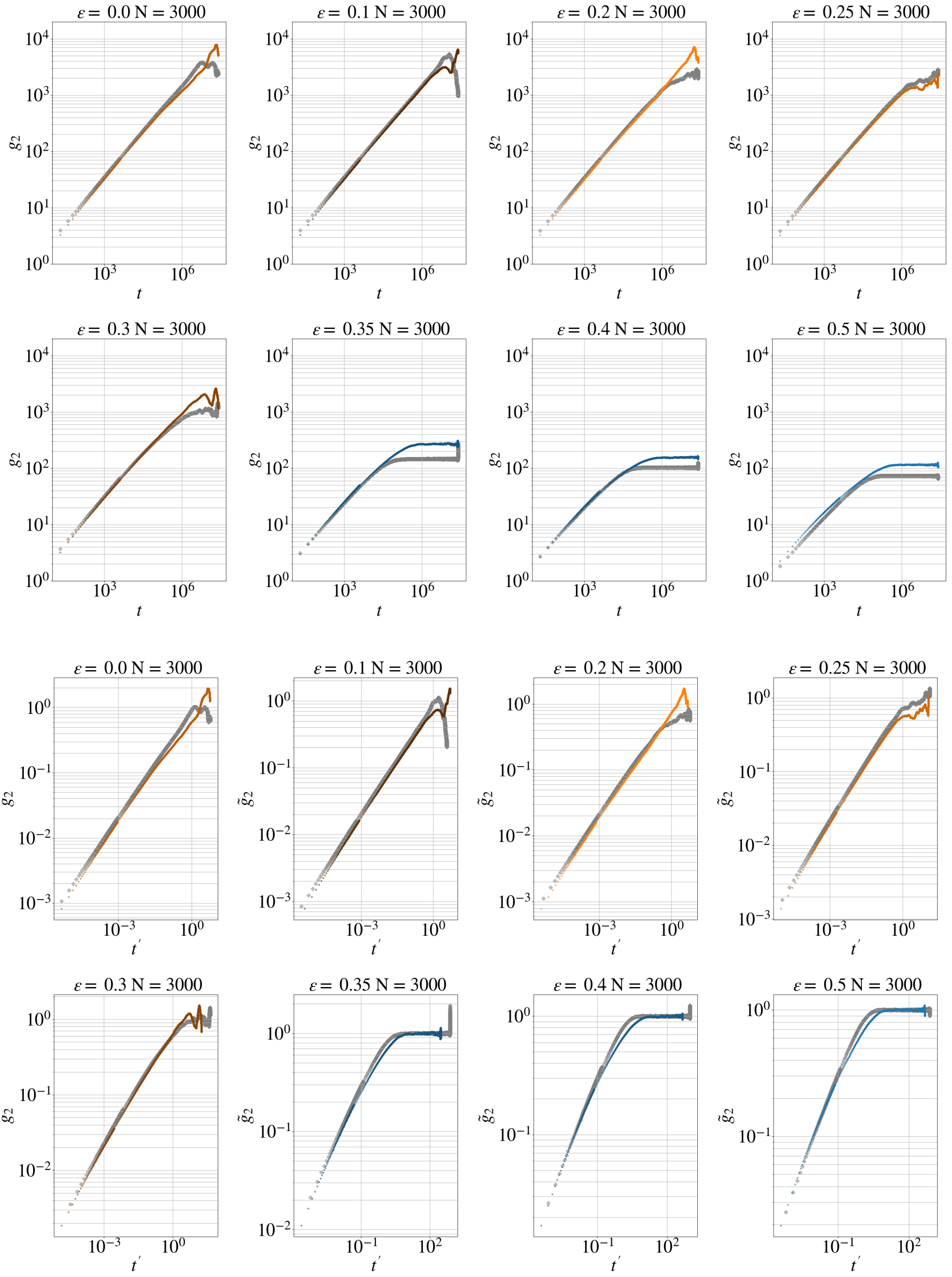


Figure 7.7:

Two top rows: $g_2(t)$ for $N = 3000$ and ϵ values all across the coil-globule phase transition.

Two bottom rows: The rescaled MSD $\tilde{g}_2(t)$ for the same simulation, such that the Rouse regimes collapse.

7.4.3 *Reptation Dynamics in the Globule Phase*

Let's therefore place our attention in the globule phase, where the largest effects due to topological constraints are expected. In Figure 7.8, $\tilde{g}_2(t')$ is plotted for $\epsilon = 0.5$ and $N = 200, 400, 800, 3000, 6000$ and 12000. For the **smallest polymer lengths** ($N = 200, 400, 800$), only a negligible deviation from Rouse behavior is observed despite the high density: **the same Rouse-like dynamics**, as in the harmonic bond simulation, is observed.

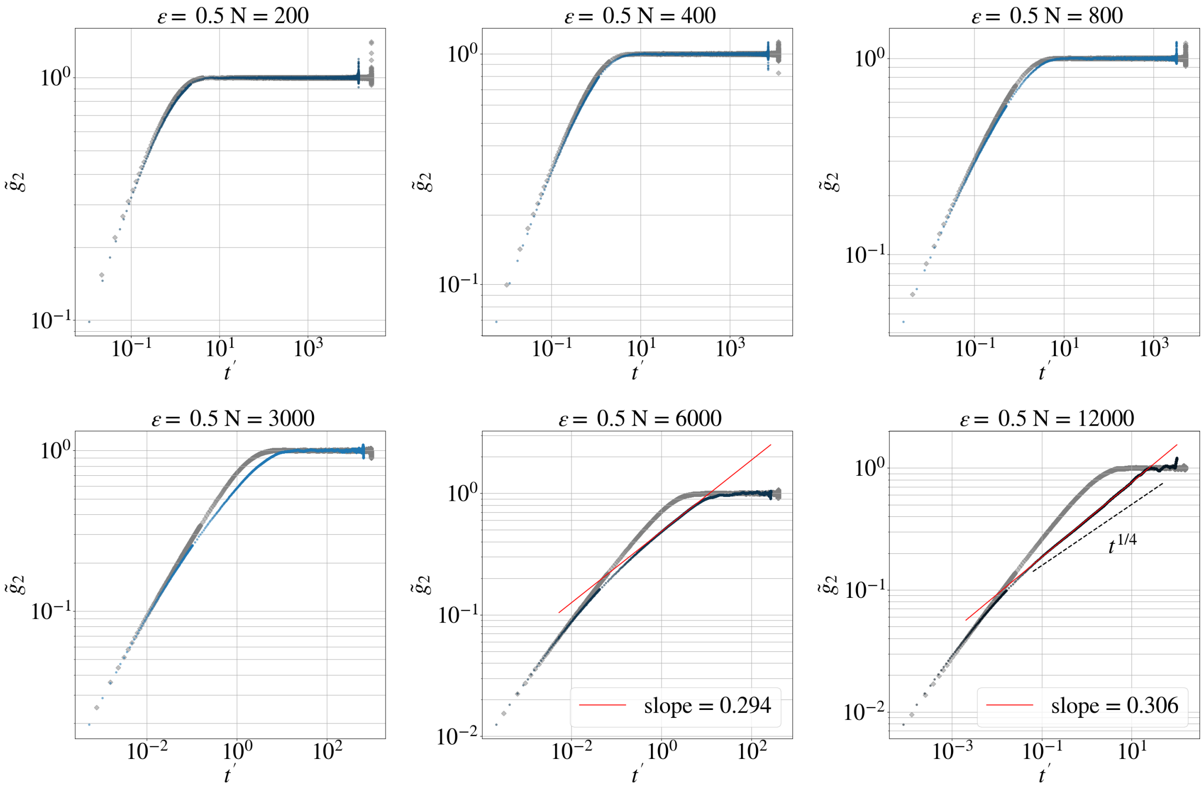


Figure 7.8: Rescaled MSD $\tilde{g}_2(t)$ for $\epsilon = 0.5$ and $N = 200, 400, 800, 3000, 6000$ and 12000.

From $N = 3000$ onward, as both g^* and M increase, a distinct regime of sub-Rousean diffusion is established, following the initial Rouse regime, marking the onset of constrained dynamics in the globule. A robust power law scaling of g_2 is observed with an exponent:

$$\boxed{\Psi_{\text{constr}} \approx 0.3.} \tag{7.10}$$

This exponent, which is found consistently in simulations for $\epsilon \geq 0.4$ and $N = 6000, 12000$, **contrasts with the** $1/4$ **exponent predicted** by tube theory for the reptation dynamics regime. This discrepancy can certainly not be attributed to the equivalent subchain size $g^* \approx 500$, which is well above N_e and is usually enough to observe proper reptation dynamics in melt simulations (the green curve in Figure 7.4). Most probably, the low number of equivalent subchains $M \approx 60$ produces a breakdown of tube theory, resulting in the observed change in the exponent. In this case, the effect is only due to a low polymer size and the correct exponent should eventually be observed for longer chain sizes. This hypothesis is further confirmed by the only other study conducted on globule dynamics[9] in which proper reptation dynamics are observed in globules of length $N = 262144$.

Following Ref. [10], an estimate of the entanglement length can be extracted from the single monomer dynamics in the globule. Indeed, the onset of reptation dynamics are marked by the entanglement time τ_e, which can be approximated as the intersection between a linear fit to the Rouse regime and a linear fit to the reptation regime. Moreover, we can assume that τ_e is the relaxation time of a Rouse chain of N_e monomers, yielding $g_2(\tau_e) \sim 2R_g^2$, and eventually

$$N_e = \frac{3D_R}{\tau_e^{1/2} b^2}, \tag{7.11}$$

where b is the average bond length and D_R the diffusion coefficient in the Rouse regime. The entanglement lengths, fitted on the longest chain at $N = 12000$, are reported in Table 7.2.

ϵ	0.5	0.45	0.4
N_e	86	105	130

Table 7.2: Entanglement lengths N_e fitted at different values of ϵ for a chain of length $N = 12000$.

This value of the entanglement length is in accordance with values found in the literature for dense melts of fully flexible linear chains[11]. It also fits well with the smallest globule in which we observed a (minute) slowing down of dynamics with respect to the harmonic simulation, $N = 800$. Indeed, the length of the equivalent independent subchains making up the $N = 800$ globule is:

$$g^* \sim 800^{2/3} \approx 86. \tag{7.12}$$

Altogether, these findings provide insights on the effects of topological constraints on the dynamics of self-interacting polymers. I will, this time, directly draw my conclusions in the final section.

[9] M. Tamm, L. Nazarov *et al.* Anomalous diffusion in fractal globules. *Physical review letters*, 114, 04 2014

[10] K. Kremer and G. S. Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086, 04 1990

[11] K. Kremer and G. S. Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086, 04 1990

7.5 CONCLUSION / TAKE-HOME MESSAGES: THE DYNAMICAL PHASE DIAGRAM OF THE FINITE SIZE COIL GLOBULE PHASE TRANSITION

7.5.1 *Topological phase diagram*

By comparing the scaling in g_2 during the constrained dynamics regime in the FENE simulation to the corresponding scaling in the harmonic potential simulations, a qualitative assessment of the onset and intensity of topological constraints on the dynamics of self-interacting polymers is depicted in Figure 7.9. The size of the red crosses is proportional to the difference in dynamic scaling exponent observed in the constrained and unconstrained case, in the reptation dynamics regime.

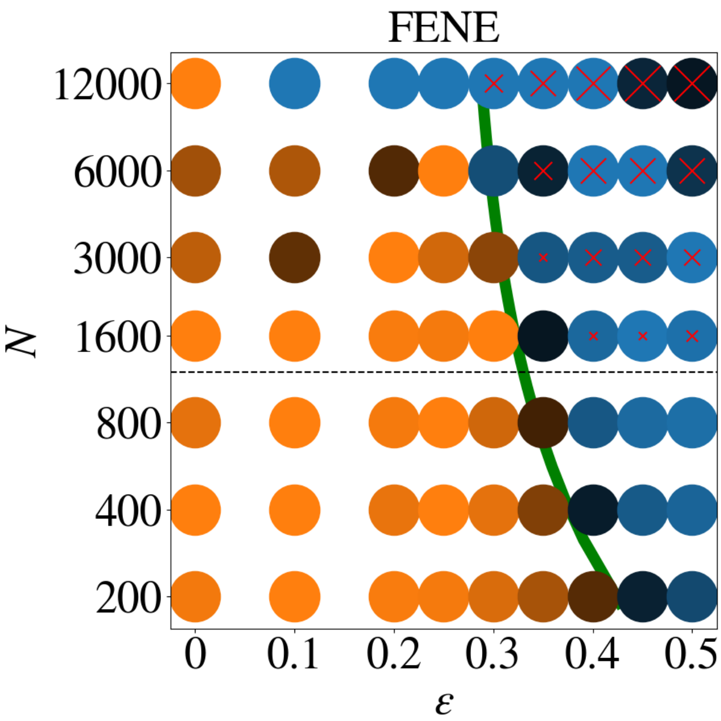


Figure 7.9: Dynamical phase diagram for the finite-size coil globule phase transition.

Impact and onset of topological constraints

- The first observation is that, as expected, topological constraints have absolutely **no impact on dynamics in the coil phase**.
- Furthermore, we observe that **below a critical length**, approximately $N = 800$, one can consider, for all practical purposes, that the **the monomer exhibits Rouse dynamics**, regardless of the strength of the folding state of the chain.

- Then, **in the globule phase, and starting from** $N = 1600$, we indeed confirm the slowing down of dynamics due to the effect of topological constraints, starting from the entanglement time τ_e, that in turn depends on the interaction energy ϵ. The effect aligns with predictions of tube theory, i.e. a **decrease of the dynamical exponent**. However, this effect is very weak for globules with size $N = 1600$ and $N = 3000$ and only becomes substantial from $N = 6000$.

- For the largest globule, $N = 12000$ a power law scaling with exponent $\Psi_{\text{constr}} = 0.3$ is observed. The larger value of this exponent compared to the 0.25 excepted from tube theory is probably due to finite size effects, since [75] indeed found an exponent of 0.25 in a larger globule.

- Moreover, looking at the dynamics close to the coil-globule transition, substantial constrained dynamics only appear once the polymer is sufficiently compacted, i.e when the polymer is well into the globule phase, in accordance with tube theory, which predicts that the entanglement length N_e increases with **density**.

- Finally, we note that, at the transition, the dynamics show no peculiarity that could have potentially appeared as a consequence of the interplay between topological constraints and the critical fluctuations related to the transition.

All in all, we find that, in collapsing polymers, Rouse dynamics are robust to topological constraints for moderate-sized polymers, $N \leq 800$. Also, it is essentially imperceptible for $N = 1600$ and $N = 3000$, even in the globule phase. Most importantly, we find that **constrained dynamics only appear inside the globule phase, meaning Rouse dynamics also survive in the transition phase**.

7.5.2 *Assessing the Coil-globule State from Dynamical Data: Finding the good observable*

Remember that our goal is to characterize the dynamics in the different phases of the model, with the idea of potentially being able to tell the folding state of a polymer from the dynamics of a single monomer. Looking back at our results for the coil and the globule phase, there are a few potential observables that could be used for such a task.

- The most robust quantity, as it is completely dimensionless, would be ψ, the **exponent of the MSD** during the internal dynamics regime.

  If the polymer is **large** and **dense** enough for topological constraints to be significant, then $\Psi < 0.5$ would be measured. Conversely, ψ stays essentially constant from the coil to the transition phase (see Figure 6.9), $\Psi \approx 0.56$, which is very close to its Rouse value. Consequently, if the polymer is either in the **coil or transition state**, the exponent Ψ would essentially assume its **Rouse value**. Hence, in the context of a self-interacting polymer, the single monomer MSD exponent Ψ can only distinguish a dense globule from a coil or transition polymer. For the effect to be measurable, the polymer should be sufficiently large, i.e. $N \gtrapprox 800$.

- Another option would involve leveraging the **drop in diffusivity** resulting from the increase in effective viscosity due to the higher packing fraction in the globule phase compared to the coil. In my simulations, for a polymer of size $N = 800$, the diffusivity was approximately divided by 2. However, the challenge here lies in the necessity for a reference diffusivity from which to compare. By determining whether we are diffusing at this rate or at a slower pace, we could potentially distinguish between the coil and globule states. Yet, establishing this reference would necessitate measuring diffusivity *in-vivo* on chromatin with a known folding state, which in itself is challenging. Moreover, a change in diffusivity could (and does) arise from a multitude of biological factors. Consequently, the most we could likely conclude is that the folding state plays a role in contributing to this change in diffusivity.
- Finally, the most promising metric is the **longest relaxation time** τ_1, that dramatically decreases as the polymer collapses. Unfortunately, its derivation from the dynamics of a **single point** is tricky if not impossible in real data. Indeed, in the simulations, τ_1 could be determined either from the MSD, taken as the time needed for the particle to diffuse the extent of the whole polymer (meaning the MSD plateaus in the center of mass reference frame), or as the crossover time between the Rouse regime and the long time whole polymer diffusive regime. This however only applies to an **isolated** chain. In a chromatin experiment, presuming the objective is to ascertain the folding state of a segment inside a whole chromosome, the "whole polymer diffusion" in this case pertains to the diffusion of the center of mass of the segment itself. Being part of a larger polymer, namely the entire chromosome, the center of mass of the segment may not exhibit normal diffusion, but rather Rouse-like diffusion due to its inclusion within a polymer. The onset of genuine whole chain diffusion would only arise at very long times when the dynamics of the whole chromosome become dominant over internal dynamics[12].

However, the relaxation time τ_1 can be determined in another way. Indeed, a common approximation for the longest relaxation time of a polymer is the **decorrelation time of the end-to-end vector** of the chain, which makes sense given the geometrical interpretation of X_1. Hence, provided the trajectory of two monomers (or two marked chromatin loci), it is possible to determine the relaxation time of the segment between the two points. Still, without a reference, with a single pair of points, it would be difficult to determine the folding state of the segment.

[12] Normal diffusion isn't even bound to ever appear, since the entirety of the chromosome is encircled by other chromosomes, which are themselves confined within the nucleus, thereby substantially limiting the dynamics of the whole polymer.

In contrast, by **varying the genomic distance between the two marked loci**, one can measure the relaxation of different-sized subchains, eventually reconstructing the relaxation spectrum τ_p for the segment. Then, assuming topological constraints don't affect the dynamics, building upon the proportionality between τ_p and $\langle X_p^2 \rangle$ (see Figure 6.4), a dynamical equivalent of the order parameter α defined in chapter 3[13], **the scaling of the long-distance relaxation times**, could as well indicate the folding state of the polymer. Then, the value of this scaling would indicate the degree of folding of the chain, with the extreme values of 0 indicating a pure globule and -2.2 a pure coil globule.

In the next chapter, based on this observation and in conjunction with the first part's conclusions concerning the critical nature of chromatin, I will propose a solution to the apparent paradox raised in [14] concerning the structure and dynamics of Drosophila chromatin.

[13] The order parameter α was defined as the slope in the first two low frequency modes in the log-log plot of $\langle X_p^2 \rangle$.

[14] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023

0

8

Assessing the Coil-globule State from Dynamical Observables

8.1 INTRODUCTION

As mentioned in the introduction to the thesis, recent papers have pointed towards an apparent paradox between two physical features of chromatin.

On the one hand, its **dynamic properties**, independently of the complexity of the organism, seem to display **Rouse dynamics**. In human cells, Keizer et al.[1] showed that the stress response of chromatin was compatible with a Rouse-like relaxation of the chain. Similarly, Bruckner et al.[2], observed Rouse-like dynamics in the pairwise motion of two distant loci in *Drosophila* cells.

On the other hand, the **structure** of chromatin, in both organisms, is known to be **more compact** than the ideal chain configuration usually associated with Rouse dynamics. In humans, this was shown through the scaling of the contact probability[3] and the sequential FISH imaging[4] (studied in chapter 4). In *Drosophila*, Bruckner et al., while measuring the dynamics, simultaneously measured the scaling of the average 3D distance between loci as a function of their genomic distance along the chain s, i.e. the internal end-to-end vector size $R(s)$, defined in equation (1.13). This scaling is found to be $R(s) \approx s^{0.31}$, which indeed deviates from the ideal chain scaling $s^{0.5}$. In both human and *Drosophila*, the measurements seem, therefore, compatible with the **fractal globule** model.

The paradox arises when considering the expected dynamics of a fractal globule chain. It can be shown, through a simple scaling argument, that the relaxation dynamics in a scale-invariant polymer model, such as the Rouse chain or fractal globule models, are completely determined by the fractal dimension of the chain[5]. In the free draining approximation, the MSD and Rouse modes relaxation times are expected to scale respectively as[6]:

$$g(t) \sim t^{\frac{2\nu}{2\nu+1}} = t^{\Psi} \quad \text{and} \quad \tau_p \sim p^{-(2\nu+1)} = p^{\alpha}. \tag{8.1}$$

[1] V. I. P. Keizer, S. Grosse-Holz *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605):489–495, 2022, [42].

[2] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023, [11].

[3] E. Lieberman-Aiden, N. L. van Berkum *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009, [49].

[4] B. Bintu, L. J. Mateo *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, 2018, [9].

[5] M. Tamm, L. Nazarov *et al.* Anomalous diffusion in fractal globules. *Physical review letters*, 114, 04 2014, [75].

[6] A. Khokhlov. *Statistical Physics of Macromolecules*. AIP series in polymers and complex materials. AIP Press, 1994

where $\nu = \frac{1}{d_F}$. For a fractal globule, $d_F = 3$, yielding:

$$g(t) \approx t^{0.4}, \qquad \tau_p \approx p^{-1.6}. \tag{8.2}$$

The expected MSD exponent for fractal globule is clearly not compatible with the observed Rouse dynamics, $g(t) \sim t^{0.5}$.

The paradox is summarized by Grosse-Holtz et al.[7] as follows:

> Within the context of commonly used polymer models for chromatin, we are thus left with two mutually contradictory observations: a fractal globule would reproduce the compact structure but with slower dynamics; the fast dynamics are consistent with the Rouse model, but that assumes an unrealistically open, equilibrium conformation. Does this point to some fundamental inconsistency in structural vs. dynamical observations, or is it simply that both models are wrong? If so, how can we reconcile all observations?

In this paper, the authors propose incorporating hydrodynamic interactions between monomers. This addition is known to accelerate the chain dynamics without changing the chain structure, offering a potential resolution for the paradox.

Further complicating the picture, Bruckner et al.[8], starting from the two-locus dynamics, could measure the relaxation times $\tau(s)$ of a chromatin sub-chain of size s, revealing the "anomalous" scaling:

$$\tau(s) \sim s^{0.7 \pm 0.2}, \tag{8.3}$$

which isn't compatible with either the Rouse scaling (s^2), or fractal globule scaling ($s^{1.6}$). This scaling indicates a **much faster** relaxation compared to both classical models.

In this chapter, utilizing elements from both parts, I'll attempt to show that, again, modeling chromatin as a critical polymer, lying at the coil-globule phase transition, accounts for both the dynamical and structural scaling, thus addressing the apparent paradox.

In section 8.2, I briefly present the experimental set-up employed by Bruckner et al. In section 8.3, I detail the results found in the original publication analysis, and there conclusions. Finally, in section 8.4, I confront the experimental data with simulated self-attracting polymers.

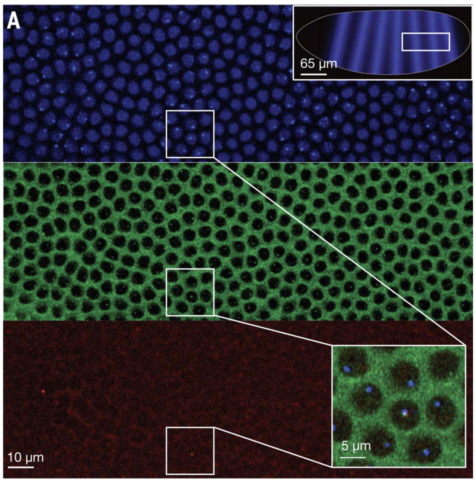


Figure 8.1: Typical view of fly cells, displaying fluorescent enhancer (blue), promoter (green) and transcript (red), from Ref. [11].

8.2 EXPERIMENTAL DATA

As mentioned in the introduction, for a gene to initiate transcription, a genomic segment, the promoter, that is adjacent to the gene, must be put into physical contact with another potentially distant genomic sequence, called the enhancer (see Figure 2). Consequently, the coupled dynamics of enhancer-promoter pairs, play a crucial role in transcription regulation and have thus been a hot topic of research in recent years. In this context, Bruckner et al.[9] investigated the effect of the

[7] S. Grosse-Holz, A. Coulon *et al.* Scale-free models of chromosome structure, dynamics, and mechanics. *bioRxiv*, 2023, [33].

[8] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023, [11].

[9] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023, [11].

genomic separation between the enhancer and promoter segments on their relative dynamics and on the transcription activity.

To that end, they developed a live-imaging setup allowing to simultaneously record the trajectories of an enhancer-promoter pair and monitoring the transcription activity of the corresponding gene[10], while systematically varying the genomic separation between the enhancer and promoter. To record the position of the enhancer (blue spot in Figure 8.2) and promoter (green spots in Figure 8.2), they placed fluorescent elements at both sites. Then the position of the promoter, and thus of the corresponding fluorescent probe, was moved along the chromatin chain. The different values of **genomic separation** s between both elements are :

$$s = 57,\ 81,\ 87,\ 148,\ 189,\ 595,\ 3327\,\text{kb}. \tag{8.4}$$

For each genomic position of the promoter, the trajectories of the enhancer and the promoter were recorded for ≈ 30 minutes, at a time resolution of $\Delta t = 30s$. These observations were performed on hundreds of cells at once situated at the posterior end of a *Drosophila* fly embryo (see Figure 8.1).

All in all, for each genomic separation s, **the resulting data** consists of a collection of hundreds of pairs of trajectories, one for the promoter, one for the enhancer (respectively $R_j(t)$ and $R_i(t)$ in Figure 8.2).

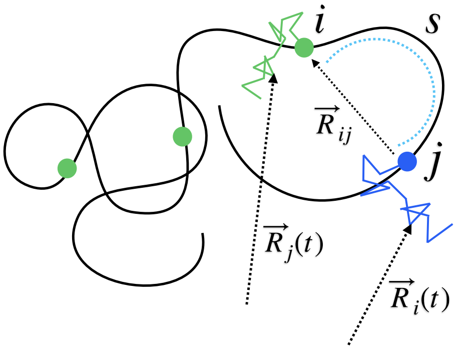


Figure 8.2: Schematic representation of the experimental set-up

8.3 DATA ANALYSIS IN THE ORIGINAL ARTICLE

From this data, the authors compute a number of structural and dynamical observables that are detailed in this section.

8.3.1 *Average End-to-end Distance Function of Genomic Separation*

To determine the inter-locus distance scaling, the authors compute the **average distance** between the two markers, $\langle R_{ij} \rangle$, as a function of their genomic distance $s = |i - j|$:

$$R(s) = \langle R_{ij} \rangle = \sqrt{\langle (\vec{R}_i - \vec{R}_j)^2 \rangle} \tag{8.5}$$

where $\langle\rangle$ represents an average over **all frames** and **all cells**, for a specific genomic seperation. The results are depicted in Figure 8.3.

The interesting data points are the blue ones (cyan and red correspond to situations where a stable loop is formed between both segments, which we won't comment on here). The dark blue dashed line indicates a power law best fit to the data for the range of **genomic separations 58 to 190 kb**, revealing the "scaling":

$$R(s) \sim s^{0.31}, \tag{8.6}$$

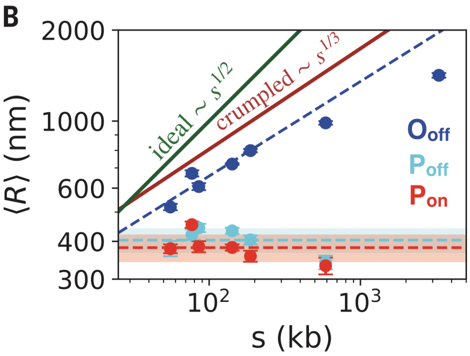


Figure 8.3: Mean inter-locus distance, function of genomic separation s. Taken from [11].

[10] In the context of this thesis, the gene activity monitoring is irrelevant, hence I won't further elaborate on it.

i.e. the one predicted for the fractal globule model. I allow myself to put ”scaling” in quotation marks, as it seems somewhat unrealistic to talk about fractality for a power-law behavior over only half an order of magnitude. For the two largest separations $s = 595, 3327$, the value $R(s)$ is lower than the $s^{0.31}$ scaling.

8.3.2 *Single Locus MSD*

To quantify the dynamics of the chain, the authors start by computing the single monomer MSD for all genomic locations, and for both the enhancer and promoter loci. These MSDs (reported in Figure 8.4, top row), are all well-fitted by a power law:

$$MSD(t) = \Gamma_1 \, t^{\beta}. \tag{8.7}$$

The fitted **single monomer diffusion coefficients** Γ_1 and exponents β are reported in the bottom row of Figure 8.4. With an average measured exponent of $\beta = 0.52$, the single monomer MSD is indeed consistent with a Rouse-like dynamics of the chain.

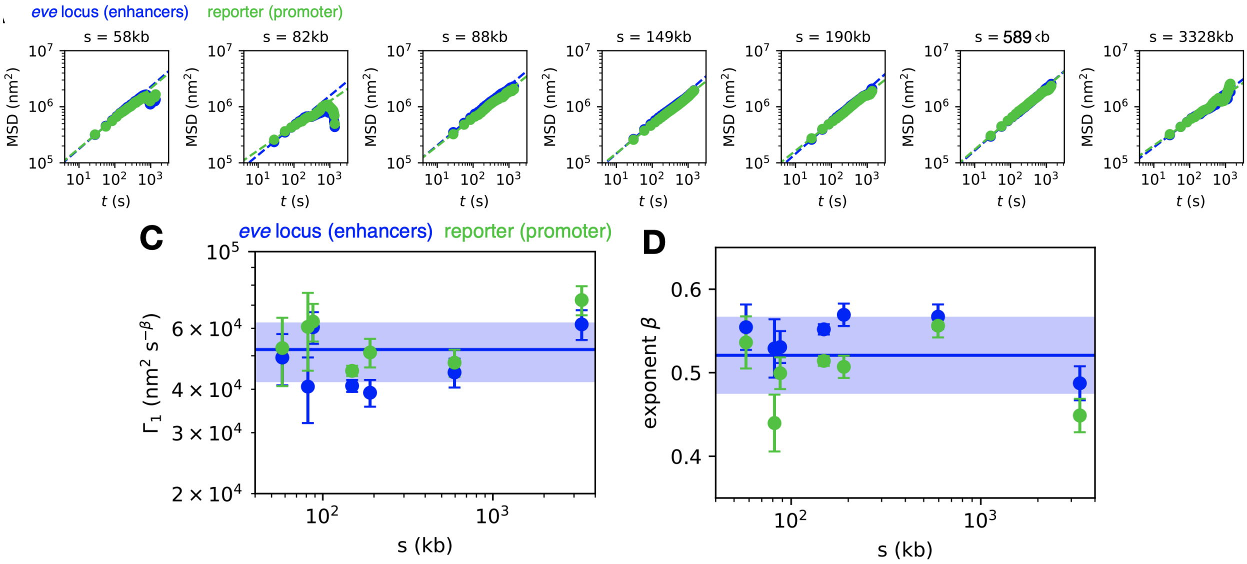


Figure 8.4: from Ref. [11]:
Top row: single locus MSD for all genomic separations for both the enhancer (blue) and promoter (green) sites.
Bottom left: fitted diffusion coefficients from a log-log linear fit.
Bottom right: fitted MSD exponents, from the same log-log linear fit.

8.3.3 *Two-locus MSD*

Interestingly, to quantify the joint dynamics, the authors compute **the two-point MSD** relative to both imaged segments, called M_2 in their paper, defined as:

$$M_2(t) = \left\langle \left(\vec{R}_{ij}(t_0 + t) - \vec{R}_{ij}(t_0) \right)^2 \right\rangle \tag{8.8}$$

where $\langle\rangle$ represents an average over time and cells. The two-locus MSD can be seen as the MSD of one locus in the reference frame where the other locus is fixed. The measured two-locus MSDs are plotted in Figure 8.5.

A theoretical expression for this quantity is available[11] for two mo-

[11] M. Gabriele, H. B. Brandão *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, 376(6592):496–501, 2022

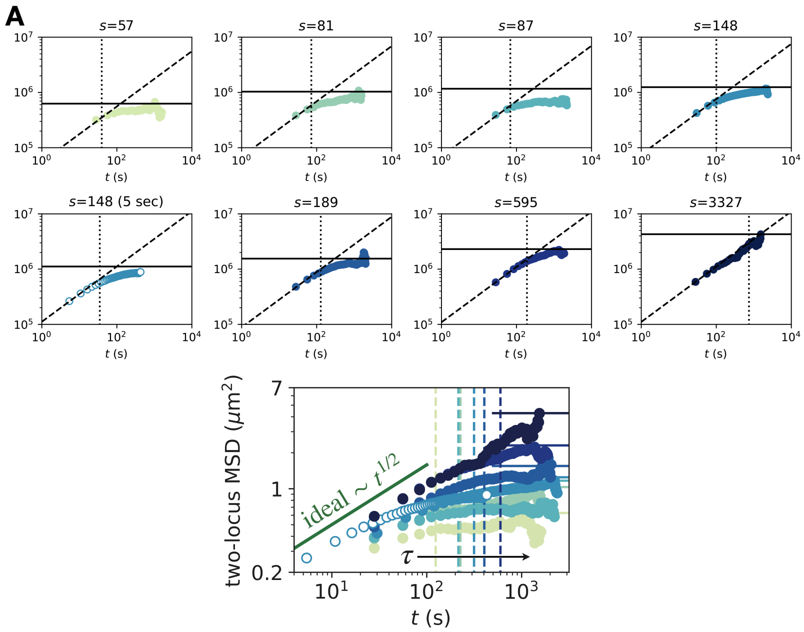


Figure 8.5: two-locus MSD for all genomic distances, from Ref. [11].

nomers, at distance s along the chain, in a Rouse polymer of size N, and in the limit $s \ll N$:

$$M_2(t)[s] = 2\Gamma\, t^{1/2}\left(1-e^{-\frac{\tau(s)}{\pi t}}\right) \;+\; 2\langle R_{ij}^2\rangle \;\mathrm{erfc}\left[\left(\frac{\tau(s)}{\pi t}\right)^{1/2}\right]. \tag{8.9}$$

Here, $\langle R_{ij}^2\rangle$ is the average squared distance between both monomers, Γ is the diffusion coefficient in the Rouse regime, defined in equation (5.70). The characteristic time $\tau(s)$ separates the two limit regimes of expression (8.9), that can be easily interpreted: At short time, $t \ll \tau$, the two-locus diffusion is determined by the independent diffusion of the two loci and is thus equivalent to the single locus MSD (modulo a factor two); at long time, $t \gg \tau$, it converges to twice the average squared inter-monomer distance, $2\langle R_{ij}^2\rangle$. The crossover between the two regimes therefore occurs at a characteristic time $\tau(s)$ given by

$$\tau(s) = \left(\frac{\langle R_{ij}^2\rangle}{\Gamma}\right)^2, \tag{8.10}$$

that is when $\Gamma t^{1/2}$ equals $\langle R_{ij}^2\rangle$. Note that, in principle, the diffusion coefficient Γ is independent of s, as we can expect it to be identical to the single monomer Rouse regime diffusion coefficient.

Bruckner and co-workers then proceed to a Bayesian fit of the data using equation (8.9), and extract the parameters Γ, $\langle R_{ij}\rangle$, from which they compute the relaxation time $\tau(s)$ from equation (8.10). The inferred Γ and τ are presented in Figure 8.6 as a function of s. They find that the diffusion coefficient seems to **increase** as a power law with the genomic separation between the two segments:

$$\Gamma \sim s^{0.27\pm 0.03}. \tag{8.11}$$

This result is, in fact, very surprising, given that the single locus diffusion coefficient is independent of the genomic location.

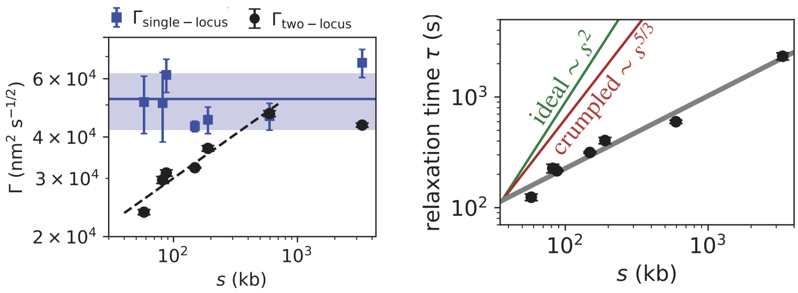


Figure 8.6: Left: single (blue) and two-locus (black) diffusion coefficients as a function of genomic separation.
Right: decorrelation times τ_s as a function of genomic separation. From Ref. [11] .

As for the segment decorrelation time $\tau(s)$, Bruckner et al. give the following power law dependence on s:

$$\tau(s) \sim s^{0.7\pm0.2}, \tag{8.12}$$

The authors compare this scaling with the expected fractal globule and Rouse scaling (equation (8.2)), respectively $\tau(s) \sim s^{1.66}$ and $\tau(s) \sim s^2$, concluding that the dynamics are substantially faster than expected for both classical models.

8.4 SIMULATION-BASED ANALYSIS SUGGESTS CRITICAL CHROMATIN ORGANIZATION

In this section, I compare the spatial structure and dynamics of simulated self-attracting polymers to this experimental data.

8.4.1 *Inter-locus Distance* $R(s)$

Let's start by modeling the inter-locus distance measured in the experiments. The correct polymer physics observable to compare with is of course the internal end-to-end distance $R(s)$ defined in (1.12).

In subsection 3.3.1 and subsection 3.4.1, we studied the scaling of $R(s)$ for both the coil and globule, which I remind here:

$$\text{Coil} \quad : \quad R(s) \sim s^{0.6}$$

$$\text{Globule} \quad : \quad R(s) \sim \begin{cases} s & \text{for } s < N^{2/3} \\ R_g^2 \sim s^0 & \text{for } s > N^{2/3}. \end{cases}$$

Hence, as ϵ increases and the polymer collapses, in the large s regime $R(s)$ goes from a power law scaling $s^{0.6}$ (coil) to s^0 (globule). Consequently, in the critical regime, the large s scaling of $R(s)$ will continuously evolve between both extreme values, and assume all intermediate values. This is depicted in Figure 8.7, where $R(s)$ for different values of N and ϵ, going through the phase transition, are depicted. Notice that, as N increases, the susceptibility of the system increases, making the range of ϵ values for which the polymer is critical progressively smaller. For this reason, it becomes challenging to simulate polymers exactly at the phase transition, for large N.

Now, remember that, in chapter 4, we found that spectral scaling properties were compatible with both the fractal globule and the critical polymer. Here, similarly, the observed scaling $\langle R_{ij}\rangle(s) \sim s^{0.31}$, which can be interpreted as a fractal globule-like arrangement, might well be the characteristic large-scale organization of a critical polymer. To back this claim, in Figure 8.7, I overlaid the experimental inter-locus-distance with the $R(s)$ for interacting polymers for $N = 1600$, 3000, 6000 at different stages of collapse.

To fit the experimental points, I scale, by hand, s and $R(s)$ in my simulations. This is equivalent to fixing two physical scales: Scaling s fixes the persistence length, b_{kb}, of the chain, i.e. the number of kilo base-pairs represented by one monomer; Scaling $R(s)$ fixes the physical size of one monomer, b.

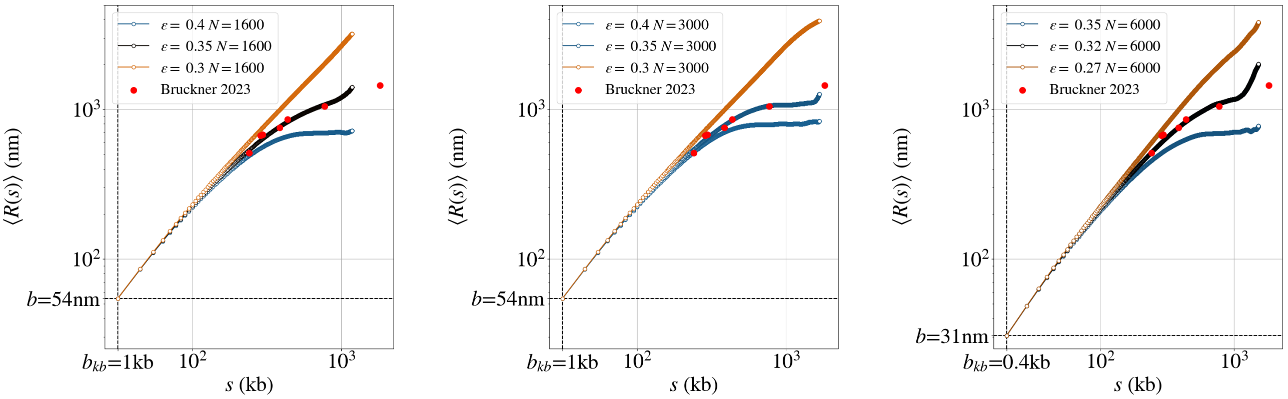


Figure 8.7: Experimental average inter-locus distance (red dots) plotted together with the internal end-to-end distance for a set of different conditions.

For all three simulated polymer lengths N, a good fit was found, albeit for different values of the interaction parameter ϵ. The adjusted values of b, b_{kb} and ϵ are reported in Table 8.1. Interestingly, for all three values of N, very reasonable adjustments of the experimental points can be obtained with **identical physical scales**, which are, moreover, on par with values used in previous polymer simulation studies of chromatin [12,13,14]. Due to the collective nature of the transition, for a fixed value of ϵ, the level of compaction for each N is different, indicated, as usual, by the color of the curves (blue for globule, orange for coil, black for transition). Hence, for $N = 3000$ and 6000, the best fit is found for a low-density globule, while for $N = 1600$ the best fit corresponds to a polymer exactly at the transition. Note, however, that it seems that also for $N = 3000$ and 6000, a critical polymer would fit better the data. However, as mentioned above, for these values of epsilon, the critical point is challenging to simulate. I'm in the process of refining these fits, by performing additional simulations.

In any case, the 5 data points at the lowest genomic separation, that exhibit the $s^{0.31}$ scaling, are well reproduced by simulated polymers close to the coil-globule transition. The conclusion we can draw is that this $R(s)$ "scaling law", attributed by Bruckner et al. to a fractal globule organization, can also be attributed to a **cross-over between two scaling regimes in a critical polymer / low-density globule**.

---

[12] A. Lesage, V. Dahirel *et al.* Polymer coil–globule phase transition is a universal folding principle of drosophila epigenetic domains. *Epigenetics & Chromatin*, 12(1), May 2019

[13] M. Socol, R. Wang *et al.* Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. *Nucleic Acids Research*, 47(12):6195–6207, May 2019

[14] A. N. Boettiger, B. Bintu *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586):418–422, 01 2016

N	1600	3000	6000
ϵ	0.35	0.35	0.32
b (nm)	54	54	31
b_{kb} (kb)	1	1	0.4

Table 8.1: Parameters used to adjust simulation and experimental $R(s)$, see Figure 8.7.

8.4.2 *Single Locus MSD*

Now that we have identified candidate simulations that match the structural aspects, let's investigate if the dynamics of these simulations are compatible with the experimental data.

In chapter 7, we've established that the dynamics of self-attracting polymers remained Rouse-like in the coil and transition phases, while reptation dynamics appeared in the globule phase, with its effect increasing with polymer length.

In Figure 8.8, the MSD for the same (N, ϵ) simulations as in Figure 8.7 are displayed. To set the simulation **time-scale** to match the experiments, I scaled time such that the experimental MSD[15] (red dash-dotted line) aligned with the simulation that fitted the best $R(s)$.

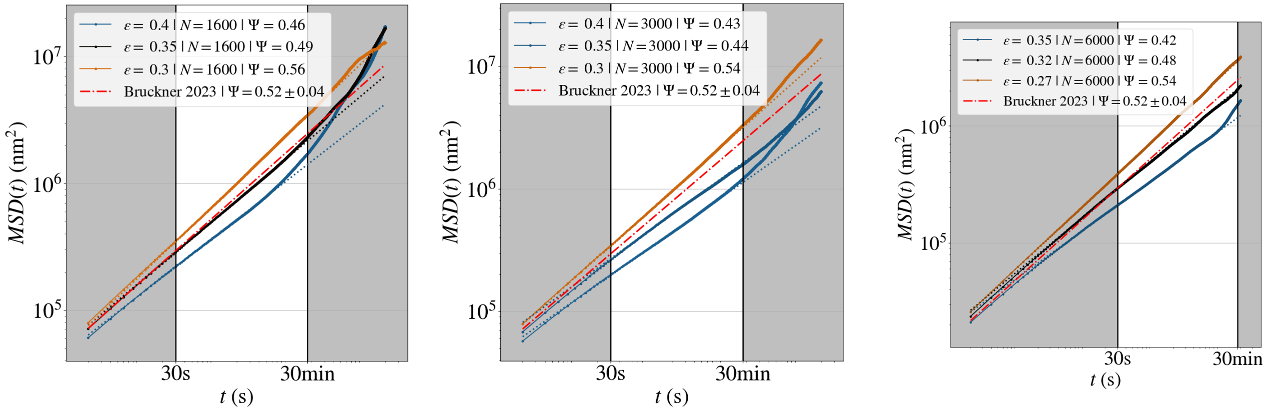


Figure 8.8: Average experimental single locus MSD inter-locus distance (red dots) plotted together with the internal end-to-end distance for a set of different conditions.

Fixing the time scale allows identifying, in the simulations, the **experimental observation time**, in between the time resolution $t = 30$ s and the total capture time $t = 30$ min. The experimental observation time is marked as a non-shaded region in Figure 8.8. Consequently, by extracting the MSD scaling in this time frame, a precise comparison of the dynamics can be conducted. The resulting MSD scaling exponents are reported in Table 8.2.

(N, ϵ)	$(1600, 0.35)$	$(3000, 0.35)$	$(6000, 0.32)$	Bruckner 2023
Ψ	0.49	0.45	0.48	0.52 ± 0.04

Table 8.2: Dynamical exponent deduced by the comparison of simulation and experiments, see Figure 8.8.

For all three sizes, a slightly sub-Rousean dynamics is observed due to topological constraints, with the strongest effect appearing in the denser polymer $(N = 3000, \epsilon = 0.35)$. Nevertheless, the effect is minute for $(N = 1600, \epsilon = 0.35)$ and $(N = 6000, \epsilon = 0.32)$, resulting in the measured exponents falling within the experimental error bars. Conversely, for $(N = 3000, \epsilon = 0.35)$, the exponent falls out of the experimental error bars.

---

[15] I made the assumption of a perfect power law Γt^{β} for the experimental MSD, parametrized by the average exponent and diffusion coefficient provided by Bruckner et al. $\Psi = 0.52$ and

In summary, the single locus dynamics of this particular chromatin segment appear to be **consistent with a self-interacting critical organization**, while a globule-like organization seems to predict dynamics that are inconsistent with the observed behavior.

8.4.3 *Relaxation Time Scaling*

Let's move on to the analysis of the two-point dynamics.

→ *Scaling Analysis*

First, it is important to note that the **expected scaling** for $\tau(s)$, given the measured structure and dynamics, is neither $\tau \sim s^2$ nor $\tau \sim s^{1.66}$, the fractal globule and Rouse model scalings, used for comparisons in the original analysis. Indeed, the fractal globule shares the same structure as measured in the experiments, but with slower relaxation dynamics. Conversely, the Rouse model shares the same dynamics but exhibits a more extended structure.

Now let's compute the expected scaling for $\tau(s)$ **given the measured structure and dynamics**. We will consider two distinct cases:

1. Considering diffusion coefficient is independent of s.
2. Taking into account the anomalous scaling of the two-point diffusion coefficient $\Gamma(s) \sim s^{0.27}$

case 1

Following the same scaling argument as in subsection 6.4.2, we can easily compute the scaling of $\tau(s)$, given $R(s) \approx s^{0.3}$ and Rouse dynamics, $MSD(t) \sim t^{1/2}$, and **without the anomalous scaling of the two-point diffusion coefficient**.

At the time $\tau(s)$ the single locus should have displaced the distance $R(s)$, yielding:

$$\mathrm{MSD}(\tau(s)) \sim \tau(s)^{1/2} \sim s^{0.62} \tag{8.13}$$

and finally:

$$\boxed{\approx \tau(s) \sim s^{1.2}.} \tag{8.14}$$

Hence a scaling analysis based on the observed $R(s)$ and MSD, already predicts a lower scaling exponent for $\tau(s)$ compared to the fractal globule and Rouse models.

case 2

Now lets take into account the s **dependent two locus diffusion coefficient** $\Gamma(s) \sim s^{0.27}$, as measured in [11]. In this case, the $\tau(s)$ scaling becomes:

$$MSD(\tau(s)) \sim \Gamma(s)\tau(s)^{1/2} \sim s^{0.62} \implies \tau(s) \sim \left(\frac{s^{0.62}}{s^{0.27}}\right)^2, \tag{8.15}$$

yielding

$$\boxed{\tau(s) \approx s^{0.7}}, \tag{8.16}$$

recovering the anomalous scaling behavior observed in the experiments.

→ *Simulation Results*

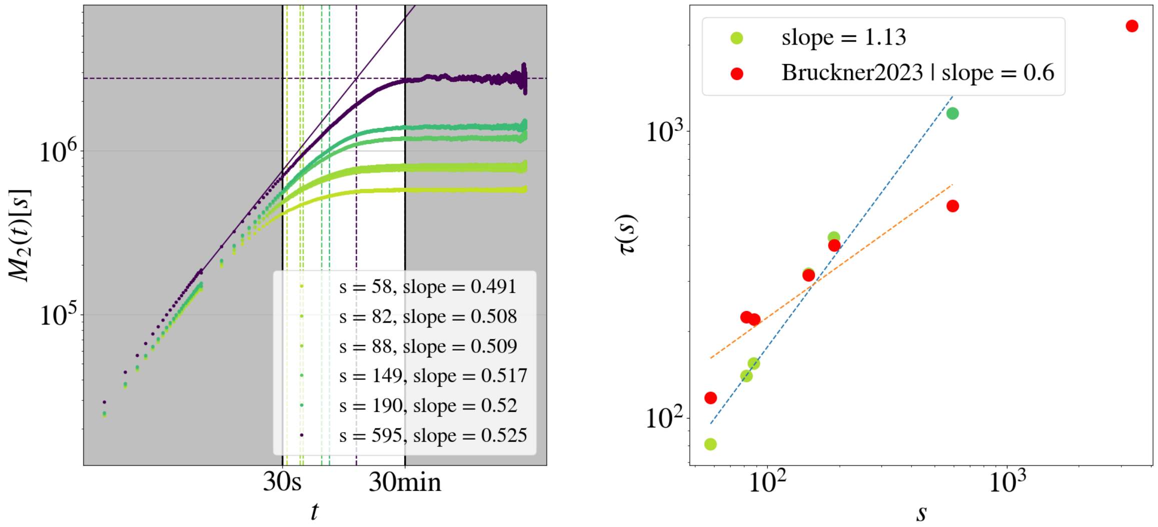


Figure 8.9: Left: two-point MSD for linear separations $s = 58, 82, 88, 149, 190, 595$ in a simulated polymer of length $N = 1600$ and $\epsilon = 0.35$. The non-shaded regions correspond to experimental observation times. The vertical dashed lines correspond to the fitted $\tau(s)$, i.e. the intercept between the asymptote $y = 2R(s)$ and a linear log-log fit in the short time regime.
Right: $\tau(s)$ from Ref. [11] (red dots) versus the fitted $\tau(s)$ (green dots). The dashed lines correspond to linear log-log fit.

Of course, in our simulations, we don't expect the two-point diffusivity to be any different from the single-monomer diffusivity, which is kept constant in the simulations. Hence, **the two-point diffusivity isn't expected to scale** with the linear separation between beads.

To check the outcome for simulations, I've nonetheless computed the two-point MSD and extracted the decorrelation times $\tau(s)$ for comparison to the experimental values. I computed $\tau(s)$ as the intercept between the linear log-log fit to the Rouse regime and the asymptote $y = 2R(s)$ (see Figure 8.9).

The results are presented in Figure 8.9 (Right) along with the two-locus MSDs (Left). I only present the analysis for $(N = 1600, \epsilon = 0.35)$, but the results are essentially identical for $(N = 6000, \epsilon = 0.32)$.

The relaxation times in the critical polymer are found to scale approximately as predicted by our scaling argument given in equation (8.14) and gives indeed

$$\tau(s) \sim s^{1.13}. \tag{8.17}$$

Hence, the **critical polymer** reproduces the $\tau(s)$ scaling expected from the experimentally measured $R(s)$ and MSD, albeit **without** the anomalous s dependence of the two-point diffusion coefficient (equation (8.14)). It consequently presents a much shallower scaling with genomic separation than expected for both the fractal globule and Rouse models. However, the anomalous two-point diffusion coefficient scaling warrants further investigation. I will discuss possible interpretations and simulation schemes to test them in the final conclusions (chapter 9)

8.5 CONCLUSIONS

In this chapter, I've compared the experimental data from Bruckner et al.[16] to simulated self-attracting polymers at different degrees of compaction.

Based on the internal distance scaling $R(s)$, I found that the population-averaged structure of the studied chromatin segment was compatible with either a low-density globule, or a critical polymer.

Moving on to dynamical observables, I've started by considering the single monomer MSD. The analysis showed that the simulated low-density globule exhibited substantially slower dynamics than the Rousean dynamics observed in the experimental data, proving it inadequate for accurately modeling the chromatin segment under investigation. Conversely, in the critical state, the polymer, while exhibiting a dense structure, indeed displayed Rousean dynamics, as observed in the experimental data.

Finally, we looked at the relaxation times $\tau(s)$ of subchains of increasing length s. The relaxation scaling in our critical polymer model is:

$$\tau(s) \sim s^{1.13} \tag{8.18}$$

which aligns with our scaling argument, given the $R(s) \sim s^{0.3}$ and $MSD(t) \sim t^{1/2}$. Hence the critical polymer exhibits a much faster decorrelation process as the fractal globule, for which $\tau(s) \sim s^{1.66}$ is expected. However, we didn't retrieve the anomalous scaling measured in the experiments, which is due to the very peculiar fact that the two-point diffusion scales with genomic distance, our model did exhibit the $\tau(s)$. In the final conclusion, I will discuss possible explanations and simple simulation schemes to test them.

As stated in the original paper, the combination of a compact structure and a 1/2 diffusion scaling exponent presents a biological advantage, allowing the search process time for distant enhancer-promoter segments to happen as fast as possible. Following this line of reasoning, it is likely that evolution has selected a critical organization of chromatin, which, as we've shown in previous chapters, is the most compact achievable form for a polymer that avoids slowing down of dynamics due to topological constraints.

Of course, the model used in my simulations has some important limitations. It's worth noting, indeed, that we've considered a homogeneous attraction between beads in our model, and an isolated chain. However, the chromatin region of interest likely contains several TADs (the median length of a TAD being around 100kb), and is flanked by the rest of the chromosome, which also involves inhomogeneities. The TADs could potentially be organized in different collapsing states depending on their epigenetic determinants and functional state[17]. It is possible that certain sub-regions of the domain under study are more compact and potentially in a more globular organization. In fact, if we take a closer look at the β exponent values experimentally determined from the single locus MSD (see Figure 8.4), for some specific cases, e.g. for s=3130 and s=82, they are below 0.5, and compatible with a low-

[16] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023, [11].

[17] D. Jost, P. Carrivain *et al.* Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*, 42(15):9553–9561, 08 2014

density globular organization. These effects may have an impact even on the averaged level and merit of course further investigation; the development of new analysis methods specifically dedicated to these effects may be a necessary step.

9

Final Conclusions

9.1 SUMMARY AND TAKE-HOME MESSAGES

Experimental evidence suggests that chromatin has a highly complex structure, capable of fine-tuning gene expression. While increasingly sophisticated experimental methods enable us to explore its characteristics, modeling its physical properties is necessary to interpret the data correctly.

Chromatin conformation capture methods have confirmed that the scaling of **population averaged** structural observables, contact probability, internal distance, and TAD gyration radius, consistently display power law behaviors at sub-chromosomal scales. The dominant interpretation for these power laws and their exponents, perhaps due to the inertia generated by its initial popularity, remains the organization of chromatin into a crumpled globule. However, polymer models, relying on radically different fundamental principles, were shown to display similar behavior on the population-averaged level[1]. Among these, it was proposed that chromatin, due to attractive interactions mediated by inter-nucleosome interactions, **lies at the coil-globule phase transition**.

In this thesis, I've developed innovative statistical methods to **test this hypothesis**. In chapters one and two, I proposed an original interpretation of the Rouse mode fluctuations as a power spectral density of the polymer conformation. In chapter three, based on extensive simulations of self-attracting polymers, I put forward a simple spectral criterion, based on the low-frequency Rouse modes, that allowed for the assessment of a polymer's folding state.

In chapter four, I proceeded to a spectral analysis of human chromatin tracing data by sequential-FISH imaging. When TAD structures are deleted by CTCF depletion, the average spectral content of chromatin conformations is consistent with the hypothesis of a critical organization. Nonetheless, the same spectral features would be observed in fractal globule organization, preventing a definitive conclusion at this stage. However, I showed that, at the single conformation level, the critical polymer was in a partially collapsed state, with nucleation sites of collapsed subchains stochastically appearing all across the chain. This property, which is yet to be observed in the

[1] M. Barbieri, M. Chotalia *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences*, 109(40):16173–16178, 2012

fractal globule, offers a physically grounded explanation for fluctuating TAD-like structures appearing in single cells after cohesin depletion.

Finally, in collaboration with Michaël Liefsoens, we developed a spectral-based loop detection method. Through this method, we could detect multiple loops in data where the conventional methods of loop detection via distance maps failed to detect their presence.

However, from the results of this part, it appears that proposing a model to explain chromatin structure cannot be done by considering only its population averaged properties, and that **dynamic aspects must be taken into account** to move the discussion forward.

In the second part, in an attempt to characterize the dynamics across the coil-globule phase transition, I carried out an **extensive simulation study of self-attracting polymers**. After recalling elements of polymer dynamics in chapter five, I analyzed in chapter six the dynamics of self-attracting polymers without topological constraints. I showed that in this simplified case, the polymer exhibited Rouse-like dynamics regardless of its folding state. In chapter seven, I added topological constraints to the simulations. By comparison to the unconstrained simulations, I could estimate the limiting conditions in which constrained dynamics appear. The results showed that constrained dynamics appeared only in the globule phase, while in the coil and transition phases, dynamics remained Rouse-like.

To conclude the thesis, in chapter eight, I analyzed dynamical data of *Drosophila* chromatin. This data, again, displayed average internal distance scaling compatible with both the fractal globule and critical polymer organizations. However, a fractal globule organization predicts a slower decorelation dynamics than that observed in chromatin experiments. As rightly stated by Bruckner et al, slower dynamics for the chromatin fiber could potentially impede essential biological processes, such as enhancer-promoter interactions, or homologous DNA repair, that involve a search process between distant chromatin segments. Conversely, the critical model assumes the limit degree of compaction before the onset of constrained dynamics. It thus predicts a chromatin fiber that is both compact and highly mobile, as it displays the fast dynamics of an ideal chain.

9.2 Limitations and prospects

→ *Anomalous Two-locus MSD Diffusion Coefficient*

The anomalous two-locus mean square displacement (MSD) diffusion coefficient measured by Bruckner et al. warrants further investigation. Clearly, either the effect is an artifact of the fitting procedure, or it is due to **long range correlations** between the two segments, that are not mediated through the chain itself. The explanation put forward in Ref.[2] is that hydrodynamic interactions between monomers might be responsible for these long-range correlations. Hence, it would be interesting to simulate critical polymers using dissipative particle dynamics to incorporate hydrodynamic correlations between beads to test this

[2] S. Grosse-Holz, A. Coulon *et al.* Scale-free models of chromosome structure, dynamics, and mechanics. *bioRxiv*, 2023

hypothesis.

From my experience in analyzing the dynamics of critical and collapsed polymers, I can also suggest, as another possibility, a potential **stochastic rotation of the nucleus**, which the experimenters cannot correct for given the location of just two points. If the two points measured are not at the same distance from the center of the nucleus, this rotation could induce correlations in their joint dynamics. I've already looked at the rotational dynamics of polymers in the glass state; it would therefore be interesting and readily achievable to simulate a polymer with rotational noise to test this hypothesis.

→ *Copolymers*

- Clearly, higher-order chromatin structures cannot be fully modeled by a homogeneous polymer. Hence, a crucial aspect that wasn't tackled in this thesis is the case of a heterogeneous polymer. It would be very interesting to study the relaxation dynamics and spectral properties of a copolymer, bearing segments of critical chromatin along with more compact portions, potentially representing repressed TADs.

- Also, throughout the thesis, chromatin segments are modeled as **isolated chains**, implicitly assuming that adjacent chromatin domains don't affect the physics of the simulated segment. However in a recent paper[1], Abdulla and coworkers, simulated a confined self-attracting polymer **embedded** inside a **non interacting polymer**, at fixed density. They showed that varying the length of the embedding polymer could drastically change the statics and dynamics of the embedded polymer, especially in the coil and transition phases. It suggests that the physical properties of chromatin segments might be heavily influenced by their surrounding genomic landscape, more so than what the isolated chain model could predict. Future work should indeed consider this phenomenon.

→ *Stress response of a critical polymer*

In light of a recent experimental paper by Keizer et al[3] that investigated the relaxation process of chromatin following a perturbation from equilibrium, conducting simulations of a critical polymer subjected to similar stress conditions and measuring its response would provide a valuable means of further testing our hypothesis concerning the critical model for chromatin

[3] V. I. P. Keizer, S. Grosse-Holz *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605):489–495, 2022

10

Bibliography

[1] A. Z. Abdulla, M. M. C. Tortora *et al.* Topological constraints and finite-size effects in quantitative polymer models of chromatin organization. *Macromolecules*, 0(0):null, 0.

[2] C. C. Aggarwal. *Neural Networks and Deep Learning: A Textbook.* Springer International Publishing, Cham, 2018.

[3] B. Alberts, A. D. Johnson *et al.* *Molecular Biology of the Cell.* 1983.

[4] E. Alipour and J. F. Marko. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic acids research*, 40(22):11202–11212, 2012.

[5] J. A. Anderson, J. Glaser *et al.* Hoomd-blue: A python package for high-performance molecular dynamics and hard particle monte carlo simulations. *Computational Materials Science*, 173:109363, 2020.

[6] E. J. Banigan, A. A. van den Berg *et al.* Chromosome organization by one-sided and two-sided loop extrusion. *eLife*, 9:e53558, apr 2020.

[7] M. Barbieri, M. Chotalia *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences*, 109(40):16173–16178, 2012.

[8] K. Binder, J. Baschnagel *et al.* Simulation of phase transitions of single polymer chains: Recent advances. *Macromolecular Symposia*, 237(1):128–138, 2006.

[9] B. Bintu, L. J. Mateo *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, 2018.

[10] A. N. Boettiger, B. Bintu *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586):418–422, 01 2016.

[11] D. B. Brückner, H. Chen *et al.* Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023.

[12] H. Chen, M. Levo *et al.* Dynamic interplay between enhancer–promoter topology and gene activity. *Nature Genetics*, 50(9):1296–1303, July 2018.

[13] A. Chertovich and P. Kos. Crumpled globule formation during collapse of a long flexible and semiflexible polymer in poor solvent. *The Journal of chemical physics*, 141:134903, 10 2014.

[14] R. Cortini, M. Barbi *et al.* The physics of epigenetics. *Reviews of Modern Physics*, 88(2), Apr. 2016.

[15] L. Costantino, T.-H. S. Hsieh *et al.* Cohesin residency determines chromatin loop patterns. *eLife*, 9:e59889, nov 2020.

[16] J. Dayantis and J.-F. m. c. Palierne. Scaling exponents of the self-avoiding-walk-problem in three dimensions. *Phys. Rev. B*, 49:3217–3225, Feb 1994.

[17] P.-G. de Gennes. *Scaling concepts in polymer physics.* Cornel University Press, 11 1979.

[18] J. des Cloizeaux and G. Jannink. *Polymers in solution. Their modelling and structure.* Oxford university press, New York, 1991.

[19] M. Doi and S. Edwards. *The Theory of Polymer Dynamics.* International series of monographs on physics. Clarendon Press, 1988.

[20] B. Duplantier. Tricritical polymer chains in or below three dimensions. *Europhysics Letters*, 1(10):491, may 1986.

[21] S. M. Espinola, M. Götz *et al.* Cis-regulatory chromatin loops arise before tads and gene activation, and are independent of cell fate during early drosophila development. *Nature Genetics*, 53(4):477–486, 2021.

[22] A. Esposito, S. Bianco *et al.* Polymer physics reveals a combinatorial code linking 3d chromatin architecture to 1d chromatin states. *Cell Reports*, 38(13):110601, 2022.

[23] J. K. Eykelenboom and T. U. Tanaka. Zooming in on chromosome dynamics. *Cell Cycle*, 19(12):1422–1432, May 2020.

[24] F. Fatmaoui, P. Carrivain *et al.* Cryo-electron tomography and deep learning denoising reveal native chromatin landscapes of interphase nuclei. *bioRxiv*, 2022.

[25] P. Flandrin. On the spectrum of fractional Brownian motions. *IEEE Transactions on Information Theory*, 35(1):197–199, 1989.

[26] G. Fudenberg, M. Imakaev *et al.* Formation of chromosomal domains by loop extrusion. *Cell Reports*, 15(9):2038–2049, 2016.

[27] M. Gabriele, H. B. Brandão *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, 376(6592):496–501, 2022.

[28] D. Gasbarra, T. Sottinen *et al.* Gaussian bridges. In F. E. Benth, G. Di Nunno *et al.*, editors, *Stochastic Analysis and Applications*, pages 361–382, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[29] P. Grassberger. Pruned-enriched rosenbluth method: Simulations of θ polymers of chain length up to 1 000 000. *Phys. Rev. E*, 56:3682–3693, Sep 1997.

[30] P. Grassberger and R. Hegger. Simulations of three-dimensional θ polymers. *The Journal of Chemical Physics*, 102(17):6881–6899, 1995.

[31] A. Y. Grosberg and A. R. Khokhlov. *Statistical Physics of Macromolecules.* AIP series in polymers and complex materials. AIP Press, 1994.

[32] A. Y. Grosberg, S. Nechaev *et al.* The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de Physique*, 49(12):2095–2100, 1988.

[33] S. Grosse-Holz, A. Coulon *et al.* Scale-free models of chromosome structure, dynamics, and mechanics. *bioRxiv*, 2023.

[34] A. S. Hansen, I. Pustova *et al.* Ctcf and cohesin regulate chromatin loop stability with distinct dynamics. *elife*, 6:e25776, 2017.

[35] H.-P. Hsu and K. Kremer. Static and dynamic properties of large polymer melts in equilibrium. *The Journal of Chemical Physics*, 144(15):154907, 04 2016.

[36] H.-P. Hsu and K. Kremer. Static and dynamic properties of large polymer melts in equilibrium. *The Journal of Chemical Physics*, 144(15):154907, 04 2016.

[37] T. A. Hunt and B. Todd. A comparison of model linear chain molecules with constrained and flexible bond lengths under planar couette and extensional flows. *Molecular Simulation*, 35(14):1153–1167, 2009.

[38] D. Jost, P. Carrivain *et al.* Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*, 42(15):9553–9561, 08 2014.

[39] M. A. Karpinska and A. M. Oudelaar. The role of loop extrusion in enhancer-mediated gene activation. *Current Opinion in Genetics & Development*, 79:102022, 2023.

[40] T. A. Kavassalis and J. Noolandi. New view of entanglements in dense polymer systems. *Phys. Rev. Lett.*, 59:2674–2677, Dec 1987.

[41] V. I. P. Keizer, S. Grosse-Holz *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605):489–495, 2022.

[42] V. I. P. Keizer, S. Grosse-Holz *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605):489–495, 2022.

[43] A. Khokhlov. *Statistical Physics of Macromolecules.* AIP series in polymers and complex materials. AIP Press, 1994.

[44] K. Kremer and G. S. Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086, 04 1990.

[45] R. Laghaei, A. Eskandari Nasrabad *et al.* Excluded volume in the generic van der waals equation of state and the self-diffusion coefficient of the lennard-jones fluid. *The Journal of chemical physics*, 124:154502, 05 2006.

[46] L. Lee, H. Yu *et al.* SnapFISH : a computational pipeline to identify chromatin loops from multiplexed DNA FISH data. *Nature Communications*, 14(1):4873, 2023.

[47] A. Lesage, V. Dahirel *et al.* Polymer coil–globule phase transition is a universal folding principle of drosophila epigenetic domains. *Epigenetics & Chromatin*, 12(1), May 2019.

[48] A. Lesne, J. Riposo *et al.* 3d genome reconstruction from chromosomal contacts. *Nature methods*, 11(11):1141–1143, 2014.

[49] E. Lieberman-Aiden, N. L. van Berkum *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009.

[50] N. Liu, W. Y. Low *et al.* Seeing the forest through the trees: prioritising potentially functional interactions from Hi-C. *Epigenetics & Chromatin*, 14(1):41, 2021.

[51] S. Liu and Q. Wang. Studying contact stress fields caused by surface tractions with a discrete convolution and fast fourier transform algorithm. *Journal of Tribology-transactions of The Asme - J TRIBOL-TRANS ASME*, 124, 01 2002.

[52] R. G. Lyons. *Understanding Digital Signal Processing.* 2010.

[53] C. Matthey-Doret, L. Baudry *et al.* Computer vision for pattern detection in chromosome contact maps. *Nature Communications*, 11(1):5795, 2020.

[54] S. T. Milner. Unified entanglement scaling for flexible, semiflexible, and stiff polymer melts and solutions. *Macromolecules*, 53(4):1314–1325, 2020.

[55] L. A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research*, 19:37–51, 2011.

[56] E. P. Nora, A. Goloborodko *et al.* Targeted degradation of ctcf decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, 169(5):930–944.e22, 2017.

[57] E. P. Nora, B. R. Lajoie *et al.* Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381–385, 2012.

[58] D. Panja and G. T. Barkema. Rouse modes of self-avoiding flexible polymers. *The Journal of Chemical Physics*, 131(15):154903, Oct. 2009.

[59] J. R. Paulson, D. F. Hudson *et al.* Mitotic chromosomes. *Seminars in Cell & Developmental Biology*, 117:7–29, 2021.

[60] L. Peliti. *Random walks with memory*, page 73. Elsevier, Amsterdam: North Holland, 1986.

[61] M. D. Pierro, B. Zhang *et al.* Transferable model for chromosome architecture. *Proceedings of the National Academy of Sciences*, 113(43):12168–12173, 2016.

[62] H. Qian. *Fractional Brownian Motion and Fractional Gaussian Noise*, pages 22–33. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[63] S. S. Rao, M. H. Huntley *et al.* A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, 2014.

[64] A. Roayaei Ardakany, H. T. Gezer *et al.* Mustache: multi-scale detection of chromatin loops from Hi-C and micro-c maps using scale-space representation. *Genome biology*, 21:1–17, 2020.

[65] T. J. Salameh, X. Wang *et al.* A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nature communications*, 11(1):3428, 2020.

[66] H. Sato, S. Das *et al.* Imaging of DNA and RNA in living eukaryotic cells to reveal spatiotemporal dynamics of gene expression. *Annual Review of Biochemistry*, 89(1):159–187, June 2020.

[67] R. D. Schram, G. T. Barkema *et al.* On the stability of fractal globules. *The Journal of Chemical Physics*, 138(22):224901, 06 2013.

[68] G. Shi, L. Liu *et al.* Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nature Communications*, 9(1):3161, 2018.

[69] G. Slade. Self-avoiding walk, spin systems and renormalization. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475(2221):20180549, Jan. 2019.

[70] M. Socol, R. Wang *et al.* Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. *Nucleic Acids Research*, 47(12):6195–6207, May 2019.

[71] G. Strobl. *The Physics of Polymers: Concepts for Understanding Their Structures and Behavior.* Springer Books, 1996.

[72] J. R. Swedlow and T. Hirano. The making of the mitotic chromosome: Modern insights into classical questions. *Molecular Cell*, 11(3):557–569, 2003.

[73] Q. Szabo, F. Bantignies *et al.* Principles of genome folding into topologically associating domains. *Science Advances*, 5(4):eaaw1668, 2019.

[74] Q. Szabo, D. Jost *et al.* Tads are 3d structural units of higher-order chromosome organization in *Drosophila*. *Science Advances*, 4(2):eaar8082, 2018.

[75] M. Tamm, L. Nazarov *et al.* Anomalous diffusion in fractal globules. *Physical review letters*, 114, 04 2014.

[76] I. Teraoka. *Polymer Solutions, An Introduction To Physical Properties.* Wiley-Interscience, 2002.

[77] M. C. Tesi, E. J. Janse van Rensburg *et al.* Monte carlo study of the interacting self-avoiding walk model in three dimensions. *Journal of Statistical Physics*, 1996.

[78] J. Van Der Maarel. *Introduction To Biopolymer Physics.* World Scientific Publishing Company, 2007.

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

[80] T. Vettorel, S. Y. Reigh *et al.* Monte-carlo method for simulations of ring polymers in the melt. *Macromolecular Rapid Communications*, 30(4-5):345–351, 2009.

[81] T. Vogel, M. Bachmann *et al.* Freezing and collapse of flexible polymers on regular lattices in three dimensions. *Physical Review E*, 76(6), Dec. 2007.

[82] S. Wang, J.-H. Su *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, 353(6299):598–602, 2016.

[83] H. Wong, H. Marie-Nelly *et al.* A predictive computational model of the dynamic 3d interphase yeast nucleus. *Current biology*, 22(20):1881–1890, 2012.