POLYMER DYNAMICS IN CROWDED ENVIRONMENT VIA SINGLE-MOLECULE IMAGING

BY

JUAN GUAN

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Materials Science and Engineering in the Graduate College of the University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Steve Granick, Chair
Professor Kenneth Schweizer
Professor Charles Schroeder
Professor Cecilia Leal
ABSTRACT

Single-molecule optical imaging is performed of entangled chain dynamics in crowded environments, for cases of both Brownian diffusion and driven transport. The molecules of λ-DNA are tracked while entangled in networks of both crosslinked gel and entangled actin filaments, and the stochasticity of dynamics is explored by acquiring millions of data points with nm resolution. Brownian diffusion displays pronounced intermittency, long-lived pausing states interrupted by transient swift hopping that couples with large-scale chain shape fluctuation. This trend persists when λ-DNA chains are driven in a preferred direction by electric fields, the chains migrating in discrete steps with strong coupling between chain motion and shape fluctuation. These trends contrast with commonly-accepted models of the chain mobility. Selective labeling of chain ends reveals a novel mode of gel electrophoresis: one end of the chain tends to stretch and pull slack from the quiescent remainder of the chain until the other end snaps forward and the cycle repeats. Methodologies developed during the course of this study, especially automated image analysis to track chain contours and modular assembly of fragments into catenated chains to track locally-labeled fluorescent regions of chain molecules, could be adapted to other related systems to resolve fundamental questions of polymer chain dynamics at the single-molecule and single-segment level with other chain architectures. A related study on diffusion in crowded hard-sphere colloid suspensions is also summarized and
included in this thesis.
To my parents, for their love and support.
First of all, I would like to thank my advisor, Dr. Steve Granick, for being a truly supportive and inspiring mentor. Joining his lab was the best decision I have ever made in my life. I have been very fortunate in the past few years to interact on a daily basis with someone who is so smart and so generous. The things I have learned from him, big or small, will surely help me with not only my future career but also appreciating the meaning of life.

I am also deeply grateful to all the group members from the Granick lab, past and present, whom I have enjoyed working alongside. In particular, Dr. Bo Wang has generously and patiently guided me in research and in life throughout the years like a big brother. I had great pleasure discussing science with him every day. Kejia Chen has supported me unconditionally as a wonderful friend and I have also learned tremendously from closely working with her on a joint project. Changqian Yu has been a wonderful collaborator on several projects and read critically my manuscripts. Dr. Chang-Ki Min taught me a lot of instrumentation and data analysis when I started my PhD. Subha Kumar also taught me SFA and FCS at the beginning of my PhD. Boyce Tsang helped me a lot with imaging and taught me LabVIEW. Dr. Lingxiang Jiang taught me how to communicate science more effectively with beautiful images. Jie Zhang helped me with surface-coating experiments. Dr. Shung Chul Bae has read critically my manuscripts and is always a great person to ask about optics. Dr. Scott Parker helped with hardware
synchronization in electrophoresis experiments. Jing Yan helped with voltage experiments and I have also learned a lot from discussing with him.

I would like to thank Dr. Charles Schroeder who provided great insight to guide me along my project. His student Dr. Chris Brockman has helped a lot with my early DNA experiments. I also want to thank Dr. Taekjip Ha for generously letting me use the equipment from his lab. The people in Ha lab have been incredibly kind to me. Among them, Dr. Jingyi Fei has taught me a lot about biological sample preparation and has been a close friend; Dr. Ruobo Zhou provided critical help with making DNA construct; Thuy Ngo helped a lot with DNA samples; Dr. Xinghua Shi helped with surface preparation; Jichuan Zhang and Boyang Hua both helped me with using equipment in Ha lab. Dr. Zhi Qi from the Chemla lab also helped me a lot with DNA sample preparation and ligation. Dr. Zidong Wang from the Yi lab and Dr. Jeanette Nangreave at ASU have given me great advices on making DNA constructs. Dr. Jimin Yao and Dr. Lanfang Li from Nuzzo lab have helped with surface patterning experiments.

I want to also thank Dr. Ken Schweizer for his great insight on several of my projects, including some that are not included in this thesis. I have benefited a lot from discussing with him and his students Dr. Daniel Sussman and Dr. Umi Yamamoto.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2 SHAPE-COUPLED MOTION OF POLYMER CHAINS IN NETWORKS AT EQUILIBRIUM</td>
<td>10</td>
</tr>
<tr>
<td>CHAPTER 3 HOPS AND STOPS IN DRIVEN TRANSPORT OF POLYMER CHAINS THROUGH HYDROGEL</td>
<td>27</td>
</tr>
<tr>
<td>CHAPTER 4 ZIG-ZAG PATHS WHEN DNA SPEEDS THROUGH HYDROGEL</td>
<td>44</td>
</tr>
<tr>
<td>CHAPTER 5 AUTOMATED LINE TRACKING OF DNA CHAINS FOR SINGLE-MOLECULE IMAGING</td>
<td>59</td>
</tr>
<tr>
<td>CHAPTER 6 DIRECT SEGMENT TRACKING THROUGH MODULAR STITCHING TO IMAGE SUB-MOLECULAR DYNAMICS</td>
<td>78</td>
</tr>
<tr>
<td>CHAPTER 7 FICKIAN YET NON-GAUSSIAN DYNAMICS IN A SIMPLE HARD-SPHERE COLLOID SUSPENSION</td>
<td>93</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>111</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

2.1 Intermittent dynamics of shape, center-of-mass and ends . . . 11
2.2 Coupling between shape fluctuation and chain motion . . . . 14
2.3 Coupling between shape and motion in agarose network . . . 15
2.4 The kinetics and magnitude of the hops . . . . . . . . . . . . 16
2.5 Ensemble-averaged mean squared displacement for for $\lambda$-
DNA in networks . . . . . . . . . . . . . . . . . . . . . . . . . . 17
2.6 Simulated chain conformation within one network mesh . . . 18
2.7 Shape matrices for free diffusion cases . . . . . . . . . . . . 21
2.8 Representative line tracking overlaid with fluorescence images 22
2.9 Other shape descriptors show coupling between shape fluc-
tuation and chain motion . . . . . . . . . . . . . . . . . . . . . 23
2.10 Intermittent dynamics of 23 kbp in gel . . . . . . . . . . . . . 24
2.11 Shape and motion coupling for 23 kbp . . . . . . . . . . . . . 24
2.12 The waiting time distribution and hop distance distrib-
tion for 23 kbp . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
3.1 Schematic of DNA gel electrophoresis . . . . . . . . . . . . . 28
3.2 The driven transport of $\lambda$-DNA chains through agarose hy-
drogel network shows intermittent dynamics with alternating
hops and stops . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
3.3 The actual velocity of hops is faster than ensemble velocity . . 30
3.4 The dispersion in the field direction shows anomalous dynamics 32
3.5 The dwell time between hops shows a field-activated trans-
port mechanism . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
3.6 The run length and duration increases with field . . . . . . . . 36
3.7 Different energy barrier distribution gives consistent thresh-
old barrier . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
3.8 The increase of moving fraction with field is a combined
effect of decreasing waiting time and increasing run duration 38
3.9 Detection of the two states is robust across a wide range
of parameter space . . . . . . . . . . . . . . . . . . . . . . . . . . 40
4.1 Field-dependent chain orientation in DNA gel electrophoresis 46
4.2 Statistics from individual scattering events . . . . . . . . . . 48
4.3 Enhanced transverse fluctuation with increasing electric field 50
4.4 The distribution of transverse displacement $\Delta y$ spreads out with time .................................................. 52
4.5 Simulated 1D random walk trajectories match dynamics observed in experiments ........................................ 54

5.1 $\lambda$-DNA (contour length 16 $\mu$m) displays various conformations in an optical microscope ........................................ 62
5.2 Feature finding from fluorescence images of DNA chains ... 64
5.3 Line tracking through the DNA molecule. ......................... 66
5.4 Further refinement of the line identified in each single image. 68
5.5 Temporal consistency check to select valid lines ................ 71
5.6 An example of applying automated line tracking analysis. ... 73
5.7 A second example of applying automated line tracking analysis. 74

6.1 Scheme of submolecular modular labeling of DNA .............. 80
6.2 end-labeled $\lambda$-DNA migrates through 1.5 wt% agarose gel . 83
6.3 asymmetry between two ends during electrophoresis .......... 85
6.4 Hooking identified by submolecular imaging ....................... 87

7.1 The experimental system of index-matched and density-matched hard-sphere colloids .......................... 95
7.2 The distribution of ergodic breaking parameter $\xi$ .............. 98
7.3 How the displacement probability distribution scales with time 99
7.4 non-Gaussian behavior changes systematically with volume fraction ..................................................... 100
7.5 No aging or change of mobility was observed for the sample at $\phi=0.55$ .................................................. 102
7.6 Conditional displacement of successive time intervals .......... 103
7.7 Persistence and exchange time are coupled ...................... 105
7.8 Velocity autocorrelation function ................................. 106
7.9 Conditional displacement of the Nth step given the displacement of the 1st step ................................. 108
CHAPTER 1

INTRODUCTION

1.1 Objective and significance

Polymer dynamics is at the core of current research interests in polymer physics. While theories in this field have been extensively developed and predictions of dynamics have been made at the single chain and chain segment level, the experimental techniques and data analysis approaches have been less satisfactory in comparison. Traditionally, most experiments involve mechanical measurements, scattering techniques, and spectroscopic studies which are limited to probe indirect and highly ensemble-averaged bulk properties arising from the collective motion of many polymer chains. Furthermore, these measurements cannot directly track dynamics at the segmental level, which serves as a fundamental measure of polymer dynamics.

The limits of those earlier experiments have been somewhat addressed by direct imaging of large (micrometer scale) biopolymers such as DNA chains as a model polymer. Because they are long enough to accommodate many dyes along the chain contour, individual chains are bright enough to be tracked under a fluorescence microscope. Moreover, the chain size typically exceeds the optical diffraction limit (~300 nm) and therefore the chain shape can be directly visualized as a flexible DNA chain adopts various conformations under thermal fluctuations. However, the studies on dynamics have been limited to characterize ensemble-averaged center-of-mass motion through combining...
individual chain trajectories [1]. Potentially interesting features of dynamics might be lost in these highly averaged analyses such as heterogeneity between individual chains, statistically rare events, and how a chain samples different dynamic states. Additionally, although there have been interesting studies concerning chain conformation under confinements or its relaxation after deformation, the shape characterization cannot provide dynamics information at the chain segment level because it is at a highly coarse-grained level and often lacking clear physical meaning [2, 3].

To make progress, we use DNA chains as a model polymer and develop two parallel approaches to improve existing technical limitations associated with studying polymer dynamics. On the one hand, we apply the concept of statistical imaging [4] to directly follow many individual chains in a high-throughput fashion. Instead of simply combining the individual trajectories to yield an ensemble-averaged property, we focus on the chain dynamics at the single-chain level, for example, how the dynamics evolves for a given chain and how the heterogeneity varies between chains. Additionally, working at the level of individual trajectories, one can extract information such as displacement distribution and transverse motion perpendicular to the direction of driven motion that were inaccessible to conventional methods. On the other hand, we develop two complementary techniques to study segmental dynamics. The first is an automated linetracking method to analyze fluorescence images of fully-labeled chains. It follows the apparent chain contour and segmental dynamics can be subsequently measured based on the contour. The linetracking method describes chain conformation in a straightforward yet quantifiable way. It works well when a chain is relatively extended, for example, under shear or extensional flow. The second is a modular stitching approach to achieve well-controlled sparse labeling along a chain such that
desired segments can be directly followed in real time with high spatial resolution below the diffraction limit and monitored rapidly with respect to the main chain contour.

The focus of this thesis is to study polymer dynamics in crowded environments. In the presence of obstacles in a crowded environment, the thermal motion, both at equilibrium and driven transport under external forces, of polymer chains are fundamental problems that have inspired decades of research [5, 6]. The understanding of these dynamic processes gives insight into mechanical properties of polymeric materials (e.g. modulus and viscosity) and procedures in polymer processing (e.g. separation and vitrification) and therefore is clearly relevant to numerous industrial applications as well.

The experimental system consists of λ-DNA chains embedded in hydrogel networks at equilibrium or under external driving forces. Our unique approaches help us gain insight into the chain dynamics that was unexpected in classical polymer physics theories. For example, for chains embedded in networks at equilibrium, although the ensemble-averaged chain dynamics suggests reptation process as the center-of-mass mean squared displacement is consistent with classic reptation scaling law, individual chains show remarkable intermittency in motion with a strong coupling between shape fluctuation and chain dynamics. We also show a new mode of driven transport in DNA gel electrophoresis where a chain migrates in discrete steps with large asymmetry between the two chain ends which indicates the role of intrachain tension in chain migration.
1.2 Organization

In the following chapters, I will illustrate and elaborate on the aforementioned general ideas. While the subject of each chapter pursues one independent topic, these studies are organized together to provide a systematic and multifaceted view of polymer dynamics in crowded environments.

Chapter 2 studies a conceptually simple system, polymer chains embedded in networks at equilibrium. Instead of continuous reptation as would be expected, we observe intermittent chain motion and large scale chain shape fluctuations coupled to chain motion.

**Chapter 2:** Numerous diffusion studies focused on center-of-mass transport while little is known about the relevance of shape fluctuation of a diffusing object. Here, we directly image fluorescently labeled λ-DNA chains threaded through F-actin or agarose network meshes at equilibrium and track chain position and shape fluctuation in real time. Using various shape descriptors for statistical analysis, we find a strong coupling between shape fluctuation and chain transport. The shape and motion of a chain are simultaneously arrested most of the time and spontaneous large shape fluctuation transiently allows a chain to break free and achieve net transport. We demonstrate this at both individual chain and ensemble level. Tracking the chain ends reveals that the end motion is also intermittent, which can lead to the arrest of shape and chain motion. The duration of the arrested phase suggests a heterogeneous transport barrier. This novel mode of transport might be important to transport in porous media and crowded environment as shape
fluctuation is widely present in a broad class of soft materials.

Given that the chain dynamics at equilibrium is different from classical reptation, we ask whether the theoretical models based on modified equilibrium assumptions such as biased reptation can accurately describe the process of driven transport through networks. Chapter 3 focuses on characterizing transport at the single-chain level and shows that the intermittent chain dynamics at equilibrium extends to driven transport. A chain stochastically switches between moving and pausing phases instead.

**Chapter 3:** Imaging individual fluorescence-labeled λ-DNA chains threading through agarose hydrogel network meshes (98.5% water content with average mesh size ∼200 nm) when pulled by a constant electric force, we observe that chain motion is intermittent with alternating hops and stops. The stops show local rattling motion with center-of-mass position confined to the mesh size whereas the hops show burst-like dynamics which lead to directed transport through the network. Separate analysis of hops versus stops from thousands of chains provides direct kinetics and distribution information that was previously inaccessible. We show that hops are activated processes overcoming transport barrier with a characteristic force-dependent rate, with shorter dwell time between hops with larger force. The effective barriers become sparser with larger force suggesting that force shapes the energy landscape by simultaneously decreasing the overall energy barrier height and skipping a subset of barriers in a heterogeneous environment. The dispersion of displacement along force direction is greatly enhanced by force and switches from subdiffusion at low field to superdiffusion at high field.

Chapter 4 builds on the system developed in Chapter 3 but asks a different
question. We noticed that as chains are driven through networks they do not follow a straight path but rather display zig-zag like motion. We characterize the transverse motion perpendicular to the primary driven transport direction with a focus on the field-dependent transverse dispersion and how microscopic turning points along an individual trajectory give rise to the observed transverse dispersion.

**Chapter 4:** Single-molecule imaging shows zig-zag paths when electric fields drive DNA up to orders of magnitude faster than Brownian diffusion through hydrogel with which it is entangled (contour length 90 times longer than the average network mesh size of 200 nm) causing molecules to appear to diffuse normal to the field direction. The higher the driving field, the higher this transverse dispersion, shifting from subdiffusive to superdiffusive with increasing field. In analogy to common Brownian diffusion at equilibrium, we measure in this driven situation the so-called van Hove distribution in the off-axis direction. This displacement distribution, we quantify over 3 orders of magnitude of probability density, finding master curves that normalize with elapsed time. They likely reflect how moving DNA scatters off the disordered static gel through which it passes, in rough analogy to how light passes through smoke and other instances of multiple scattering.

Chapter 5 goes beyond the center-of-mass motion to study dynamics at the level of chain segments. A fully-automated linetracking algorithm is developed to follow the apparent chain contour from fluorescence images of fully-labeled chains. The segment dynamics can then be approximately followed by placing fiducial markers along the tracked contour. The method describes chain shape in a straightforward yet quantifiable way, and should be broadly applicable to chains with extended conformation.
Chapter 5: We describe a straightforward, automated line tracking method to visualize within optical resolution the contour of linear macromolecules as they rearrange shape as a function of time by Brownian diffusion and under external fields such as electrophoresis. Three sequential stages of analysis underpin this method: first, “feature finding” to discriminate signal from noise; second, “line tracking” to approximate those shapes as lines; third, “temporal consistency check” to discriminate reasonable from unreasonable fitted conformations in the time domain. The automated nature of this data analysis makes it straightforward to accumulate vast quantities of data while excluding the unreliable parts of it. We implement the analysis on fluorescence images of λ-DNA molecules in agarose gel to demonstrate its capability to produce large datasets for subsequent statistical analysis.

Chapter 6 introduces a direct segment tracking method by labeling specific segments along a chain and resolving the motion of the labeled segments in real time below the diffraction limit. This could be particularly useful to study segment dynamics when a chain is highly looped or coiled and cannot be resolved from conventional diffraction-limited fluorescence images. The method allows high design flexibility and could be conveniently adapted to numerous other polymer physics problems.

Chapter 6: For study of time-dependent conformation, all previous single-molecule imaging studies of polymer transport involve fluorescence labeling uniformly along the chain, which suffers from limited resolution due to the diffraction limit. Here we demonstrate the concept of submolecular single-molecule imaging with DNA chains assembled from DNA fragments such that a chain is labeled at designated spots with covalently-attached fluorescent dyes and the chain backbone with dyes of different color. High density
of dyes ensures good signal-to-noise ratio to localize the designated spots in real time with nanometer precision and prevents significant photobleaching for long-time tracking purposes. To demonstrate usefulness of this approach, we image electrophoretic transport of λ-DNA through agarose gels. The unexpected pattern is observed that one end of each molecule tends to stretch out in the electric field while the other end remains quiescent for some time before it snaps forward and the stretch-recoil cycle repeats. These features are neither predicted by prevailing theories of electrophoresis mechanism nor detectable by conventional whole-chain labeling methods, which demonstrate pragmatically the usefulness of modular stitching to reveal internal chain dynamics of single molecules.

Going beyond these DNA systems, we also have related studies of dynamics in crowded hard-sphere colloids, of polymers on surfaces with tunable affinity, of entangled semiflexible polymers. Chapter 7 includes surprising dynamics observed in a simple colloidal suspension.

**Chapter 7:** We scrutinize 3 decades of probability density displacement distribution in a simple colloidal suspension with hard sphere interactions. In this index-matched and density-matched solvent, fluorescent tracer nanoparticles diffuse among matrix particles that are eight times larger, at concentrations from dilute to concentrated, over times up to when the tracer diffuses a few times its size. Displacement distributions of tracers, Gaussian in pure solvent, broaden systematically with increasing obstacle density. The onset of non-Gaussian dynamics is seen in even modestly dilute suspensions which traditionally would be assumed to follow classic Gaussian expectation. The findings underscore, in agreement with recent studies of more esoteric soft matter systems, the seeming ubiquity of non-Gaussian yet Fickian diffusion.
1.3 Future prospects

With insight gained during the course of this study on the chain dynamics that was unexpected in classical polymer physics theories, there are many interesting questions one can ask which seem to be a natural extension from the current work. For example, do the phenomena we report here apply to the polymer melts? How should we think of other non-equilibrium conditions such as shear systems?

Furthermore, the methodology developed in this study should facilitate discovery in the field of polymer physics. For example, the line tracking method is broadly applicable to quantify chain conformation when a chain is extended. One can envision its convenient use in many other non-equilibrium processes. This kind of conformation information might be helpful to understand the underlying mechanism, yet inaccessible to previous image analysis methods.

Another example is the modular assembly approach for direct segment tracking. It is challenging to resolve dynamics at the segment level especially if a polymer chain is looped or coiled. The modular assembly approach allows direct tracking of chain segment with high spatial and temporal resolution in various microenvironments. A very exciting new direction we are moving into is to study segmental dynamics in more complex molecular architecture such as branched polymers.

Additionally, it has recently been brought to our attention that the way we use electric field to stretch DNA chains on surfaces might provide a high-throughput imaging platform for studying DNA-associated enzymes, which is an active research field in bioanalytical assays.
2.1 Background

Brownian motion is a fundamental phenomenon that continues to fascinate researchers. While diffusion of spherical objects has been well understood, recent studies show that shape of a diffusing object adds additional complexity to transport dynamics [7, 8]. An unexplored aspect in transport dynamics is when shape is flexible and fluctuates spontaneously. Little was known about the relevance of shape fluctuations because most previous research focused on center-of-mass dynamics without visualizing shape or lacked a feasible way to analyze shape fluctuations especially in relation to transport dynamics [1–3]. Here, we demonstrate a strong coupling between shape fluctuation of polymer chain and its transport in networks at equilibrium with direct imaging of single DNA chains. We chose DNA chains because they are big enough for shape visualization in real time and flexible enough to show spontaneous shape fluctuations. Shape information is extracted from images with description at various levels of details from pixel-based multi-dimension to coarse-grained one-dimension and correlated with the center of mass motion. The prevailing pattern is that the shape and motion of the chains are simultaneously arrested most of the time and spontaneous large chain extension or retraction transiently allows a chain to break free and achieve net transport. Tracking chain ends reveals that blocked end motion is consistent with the
arrest of shape and chain motion. This novel mode of transport might be important to transport in porous media and crowded environment as shape fluctuation is widely present in a broad class of soft materials.

2.2 Results

We embed $\lambda$-DNA chains uniformly labeled with fluorescence dyes in optically transparent network as sketched schematically in Figure 2.1a. Network

Figure 2.1: Intermittent dynamics of shape, center-of-mass and ends for chains embedded in networks. a) Schematic of a $\lambda$-DNA chain (~16 m in length) embedded in agarose or actin networks (~200 nm mesh size). The open space in the network is filled with water. The magnified view shows schematically flexible turns and bends of chain segments within one mesh. b) (top) Shape matrix $S$ calculates pixel-based correlation between pairwise images for a chain with each element indicating how similar two images are. Examples for both networks are shown. Both axes of the matrix denote time in seconds. Red blocks indicate high similarity between two images whereas blue blocks indicate low similarity. A color bar is shown for comparison. (bottom) Center-of-mass position for the corresponding chains. Top and bottom panels share the time axis. Details see main text. c) Position versus time for two ends of one chain. The two ends are often confined in different meshes microns apart.
concentration is tuned such that a chain (∼16 µm in contour length; ∼50 nm in persistence length) threads through ∼20-40 meshes (∼200 nm) to be tightly entangled with the network (supporting information). The large size of λ-DNA allows shape imaging (much exceeding the diffraction limit ∼0.3 µm). To test the generality of the finding, two types of network with matching mesh size are chosen: agarose hydrogel and semidilute solution of actin filaments. The former represents static network as crosslinked agarose fibers form interconnected meshes whereas the latter represents dynamic network as actin filaments fluctuate and rearrange locally over observation time window [9–12].

DNA chains adopt diverse shapes in these networks, from collapsed globule to extended chains with many twists and turns. Unexpectedly, in both networks the chains can keep a relatively constant shape for extended periods of time. This arrested shape dynamics is very well captured by the shape matrix $S$ (Figure 2.1b), with $S(m,n)$ calculated as the pixel-based correlation between frame $m$ and $n$ in the time series of images of one specific chain and indicates how similar these two images are. Figure 2.1b shows the shape matrices of two 62 s long movies from both networks, with red and blue representing high and low similarity respectively. Evident to the eye, the matrices are composed of blocks with different color and size. Along the diagonal line, transition from one square to the next corresponds to large change in chain position as seen from the steep slope in the position of center-of-mass plot below the matrices. The fact that the diagonal squares are mostly red indicates that the shape and position of a chain remains mostly constant within each square. The arrest of shape and chain position is really surprising because it is commonly assumed that flexible polymers in network smoothly and continuously slither back and forth in “tubes” formed by its
surrounding obstacles. A key assumption in this reptation picture is that the chain ends can freely explore the network. To test this assumption, we labeled the chain ends with fluorescence dyes and find that they are locally confined on a length scale comparable to mesh size with occasional hops (Figure 2.1c). Furthermore, the two ends seem to move and pause at different times, independent to one another. The distance between ends fluctuates at around 5 mesh size and two ends often stay in different meshes microns apart, confirming the tightly entangled regime where a chain threads through many meshes. Given the blocked ends motion and transverse confinement along the chain contour, it is reasonable to expect shape and position arrest. We note that the blocked ends motion is not an artifact from projection of vertical motion as the depth of focus is only ~1 µm and out-of-focus images are easily discerned and excluded from analysis.

It is clear from the shape matrix and the center-of-mass motion that arrested dynamics dominates in time and only the occasional swift hops are responsible for the transport of the chains. When taking a closer look at hops, we saw that a chain typically shows shape change during hops, in contrast to almost no shape fluctuations other time (Figure 2.2a). To see how general this pattern is, we used multiscale wavelet analysis [13] to automatically identify the hops across hundreds of chains and then compared shape fluctuation during these hops versus local motion. The appropriate separation between hopping and local rattling is confirmed with the distinct fast and slow center-of-mass dynamics (Figure 2.2c). Local rattling motion occupies 90% of the time determined by wavelet analysis. As extension and retraction appear to be the dominant form of chain shape fluctuation, we use apparent chain length here as a measure of global chain shape [14]. Checking across hundreds of chains, we consistently find large chain extension or
Figure 2.2: Coupling between shape fluctuation and chain motion. a) Sequential images taken from a typical center-of-mass hop, as highlighted with a red arrow in Figure 2.1b. The initial center-of-mass position is overlaid for comparison (+). The chain extension is accompanied by position hop. Image size is 2.9 $\mu$m$\times$1.5 $\mu$m. b) Chain position hop and chain length fluctuation occur simultaneously; position arrest and shape arrest occur simultaneously. Hundreds of individual traces are shown. Red curve denotes average. c) Mean squared displacement (MSD) during hops than arrest. The hops are identified from $\sim$1,000,000 images using a wavelet method. Solid curves are for agarose and dashed curves are for actin network. d) Average apparent length ($L$) change $\Delta L = \sqrt{\langle \Delta L^2 \rangle}$ during hops is larger and faster than during arrest. Symbols are the same as in c. retraction during hops whereas almost no length change is observed during local rattling (Figure 2.2b). Other shape indicators are checked in parallel to confirm the coupling between shape fluctuation and chain motion exists in
both networks (supporting information). We further compare the ensemble-averaged apparent length change ($\Delta$) with time during hopping versus local rattling from over 1,000,000 images. The shape fluctuation during hopping is significantly faster and larger than local rattling (Figure 2.2d), confirming that chain motion is accompanied by chain extension or retraction. Physically, the blocked chain ends can occasionally hop to other meshes and create entropic tension between the chain segments so the chain shape changes to release the tension and causes center-of-mass position shift.

The local confined motion suggests transport barriers imposed by the network. The waiting time distribution can be fitted with stretched exponential with exponent of $\sim 0.7$ (Figure 2.4a), suggesting a heterogeneous distribution of barrier height. The hop length has an average of $\sim 0.25 \, \mu m$ for both net-

![Graph showing hopping behavior](image)

**Figure 2.3:** (top) A magnified view of the trajectory for the $\lambda$-DNA chain in agarose network from Figure 2.1b showing a hop in chain position. (bottom) Sequential images corresponding to the hopping. The initial center-of-mass position is overlaid for comparison (+). The chain extends and retracts to a new position. Image size is $2.2 \, \mu m \times 1.7 \, \mu m$. 

15
Figure 2.4: The kinetics and magnitude of the hops. a) Waiting time between hops follows stretched exponential distribution with exponent of \( \sim 0.7 \). b) Distribution of hop distance for individual hops. Solid symbols are for agarose and open symbols are for actin network.

works, comparable to the mesh size (Figure 2.4b) suggesting that the scale of discrete hops is determined by the underlying network mesh. This length is comparable to the hop length of chain ends (supporting information) further supporting the hops are initiated at chain ends. It is worth noting that other chains sufficiently long to observe shape (e.g. half \( \lambda \)-DNA length) dis-
play similar behavior such as intermittent dynamics, coupling between chain motion and shape fluctuation, and comparable hop length.

2.3 Discussion

The ensemble-averaged chain dynamics is initially sub-diffusive with a scaling of 0.4 and 0.6 for agarose and actin network respectively and transits to diffusion at \( \sim 10 \) s, as shown in Figure 2.5. It is commonly assumed that flexible polymers in solution smoothly and continuously slither back and forth in “tubes” formed by its surrounding obstacles. It is only remarkable now

![Figure 2.5: Ensemble-averaged mean squared displacement for λ-DNA chains in actin and agarose networks.](image-url)
that direct visualization suggests alternative process involving intermittent
dynamics and shape fluctuations coupled to chain motion. Hindered motion
is unlikely to be caused by specific interactions as we demonstrate in two
distinct systems although its exact origin is unknown. Tentatively, if we con-
sider DNA chain as worm-like-chain (WLC) with persistence length on the
order of mesh size, it might have trouble wriggling through rigid network
meshes. A representative WLC chain conformation spanning the dimension
of one network mesh is shown in Figure 2.6. Taken together the intermit-
tent dynamics of shape, center-of-mass, and chain ends, it is likely that the
effective confinement experienced by a chain spans multiple interconnected
meshes hosting different parts of a chain along its contour. It remains to be
tested whether similar dynamics exists in polymer melt systems where probe
chain and surrounding chains are more flexible. Interesting questions remain
on how the phenomenon at equilibrium might transfer to non-equilibrium
dynamics where different microscopic processes compete differently. On the
practical side, the long waiting time observed here might slow down diffu-

Figure 2.6: A representative chain conformation within one network mesh
for a simulated worm-like-chain.
sion processes, an effect less studied in the context of polymer dynamics in networks, for applications in transport, mixing, and spatial distribution of constituents in networks.

2.3.1 Determine entanglement regime

We estimate the entanglement level for \(\lambda\)-DNA in both networks to be \(\sim\)20-40, a regime commonly accessed in simulations and experiments and typically assumed to represent long-chain condition. First, we estimate the entanglement level based on lengthscale of the system. \(\lambda\)-DNA (\(\sim\)16 \(\mu\)m) contains \(\sim\)160 Kuhn blobs (\(\sim\)0.1 m). Taking the average value of \(\sim\)0.20 \(\mu\)m which can contain \(\sim\)4 Kuhn blobs per mesh (\(\sqrt{4} \times 0.1 \mu m \approx 0.20 \mu m\)) and therefore the whole chain fits in \(\sim\)40 meshes. Worm-like-chain model predicts each mesh contains a chain segment with a contour length of 0.45 \(\mu\)m which would lead to \(\sim\)35 meshes for a whole chain. Next, we rule out the possibility that a chain on average occupies one single big mesh considering the heterogeneity of the networks. This is because the two ends of a chain often stay in different meshes indicating the mass is distributed in multiple network meshes otherwise the two ends should often co-localize to the same mesh that contains the entire chain. Based on the measured end-to-end distance of \(\sim\)1.5 \(\mu\)m and reported mesh size, the chain is distributed in \(\sim\)50 meshes, consistent with earlier estimate. Therefore, the effective confinement experienced by a chain is likely spanning multiple interconnected meshes hosting different parts of a chain. Third, we apply the same calculation to an artificial 2D array of \(\sim\)0.92-\(\mu\)m pores for comparison [15]. \(\lambda\)-DNA was found to mainly distribute to two pores. One pore can contain \(\sim\)80 Kuhn blobs (\(\sqrt{80} \times 0.1 \mu m \approx 0.9 \mu m\)), consistent with the current experiments.
2.3.2 Theoretical timescale estimation under reptation framework

The experimental system shows that the mean squared displacement of center-of-mass of λ-DNA transits from sub-diffusion (with an approximate scaling of $\frac{1}{2}$) to diffusion (scaling of 1). Although consistent with reptation prediction, the direct observation of microscopic process suggests an alternative explanation as discussed in Section 2.2. In reptation picture, the thermal motion of center-of-mass has a scaling exponent of $\frac{1}{2}$ before reptation time because it is a random walk within a tube of random walk configuration. The dynamics transit to diffusion with a scaling exponent of 1 beyond reptation time. The reptation time is estimated to be $\sim 16$ s for this system. Each entanglement blob has a relaxation timescale of 2 ms assuming the size of each blob ($b$) is 0.2 $\mu$m, close to pore size of the network ($\tau_0 = \frac{\eta_{\text{water}} b^3}{kT} = 2$ ms). Reptation time $\tau_R = N^3 \tau_0 = 16$ s assuming a chain contains $\sim$20 entanglement blobs. For consistence check, we also estimate Rouse time of the current system as 0.8 s ($N^2 \tau_0$). This value is consistent with literature values when additional local friction from the matrices is considered compared to free solution [16–18].

2.3.3 On shape matrix S

The individual elements of shape matrix $S(m, n)$ are calculated as the pixel-based Pearson correlation coefficient between frame $m$ and $n$ in the time series of images of one specific chain and indicates how similar these two images are. Both axes of the matrix represent time. By definition, the matrix is symmetric and the value of diagonal elements is 1. High value (red) indicates high similarity between the two images whereas low (blue) value indicates low similarity. Block feature of shape matrix is unique to
Figure 2.7: (top) shape matrices in the case of free diffusion of λ-DNA in sucrose. (bottom) shape matrices for simulated Brownian diffusion of λ-DNA in dilution solution.

intermittent dynamics in networks. The blocks suggest that chain shape and position is arrested for an extended period of time. A few representative shape matrices in dilute solution case and simulated Brownian motion case are shown in Figure 2.7 for comparison. Block feature is absent.
2.3.4 Various shape descriptors show consistently a strong coupling between shape fluctuation and chain motion

Apart from apparent length we report earlier (examples are shown in Figure 2.8), we also used other shape descriptors such as radius of gyration, long and short axis, and the shape index (Figure 2.9).

2.3.5 Shorter chains display similar intermittent dynamics and transient shape fluctuations during hops

Chains of 23 kbp, about half the λ-DNA length, is sufficiently big for visualizing shape. Figure 2.10 shows similarity matrix for a representative chain. The block pattern is similar to the λ-DNA system showing intermittent dynamics. Figure 2.11 shows snapshots from a hopping event where transient extension and retraction is coupled to chain motion. Figure 2.12 shows that both the distribution of waiting time between hops and the hop distance distribution are similar to the case of λ-DNA.

Figure 2.8: Representative line tracking to measure apparent contour length change with tracked lines (red) overlaid with raw images taken from Figure 2.2. The ensemble-averaged apparent length change calculated uses this equation: \[ \Delta L = \sqrt{\Delta L^2(\tau)} = \sqrt{\frac{1}{\tau} \int_0^{\tau} [L(t+\tau) - L(t)]^2 dt} \]
Figure 2.9: Other shape descriptors show coupling between shape fluctuation and chain motion. On the right is one sample trace to show we identify center-of-mass hop from wavelet analysis ($C_x$). Within the same time window, the wavelet coefficient of shape index ($C_s$) shows a peak which indicates big change in shape. Here, shape index is taken as the elements from the first row of the shape matrix $S$. On the left, we also show radius of gyration change coincides with chain hopping for both networks. Hundreds of individual chains are shown in gray, and the average are shown in red.

### 2.4 Conclusions

Numerous diffusion studies focused on center-of-mass transport while little is known about the relevance of shape fluctuation of a diffusing object. Here, we directly image fluorescently labeled λ-DNA chains threaded through F-actin or agarose network meshes at equilibrium and track chain position and shape fluctuation in real time. Using various shape descriptors for statistical analysis, we find a strong coupling between shape fluctuation and chain
transport. The shape and motion of a chain are simultaneously arrested most of the time and spontaneous large shape fluctuation transiently allows a chain to break free and achieve net transport. We demonstrate this at both individual chain and ensemble level. Tracking the chain ends reveals that the end motion is also intermittent, which can lead to the arrest of shape and chain motion. The duration of the arrested phase suggests a heterogeneous transport barrier. This novel mode of transport might be important.

Figure 2.10: On the left is a representative center-of-mass trajectory of 23 kbp in gel. Trajectory lasts ~50 s. The corresponding shape matrix is shown on the right. Three diagonal blocks reflects three main localizations in the trajectory.

Figure 2.11: Sequential images from a representative hopping event for a DNA chain of 23 kbp in gel. The duration of the event is 1 s. Image size is 2 µm × 1.3 µm.
Figure 2.12: The waiting time distribution between hops (left) and hop distance distribution (right) for the DNA sample of 23 kbp (half of λ-DNA length) in agarose network. The average waiting time is 4.7 s, and the average hop distance is 0.33 µm.

to transport in porous media and crowded environment as shape fluctuation is widely present in a broad class of soft materials.

2.5 Materials and Methods

λ-DNA chains (48 kbp, 16 µm in length) are covalently labeled with dyes (Mirus Bio) uniformly along the contour. Covalent attachment eliminates the extension of contour length that results from using conventional intercalated dyes. Chains were embedded in 1.5 wt% agarose gel (Fisher, molecular biology grade, low EEO) in the presence of 1× TBE and anti-photobleaching buffer. Alternatively, chains were embedded in F-actin solution (2 mg/ml) polymerized in vitro in F-actin buffer with an ionic strength of 100 mM in the presence of anti-photobleaching buffer. Both networks have a mesh size of ~200 nm. The DNA chains are visualized using a homebuilt epifluorescence microscope setup (Zeiss observer.Z1, 100× objective with N.A.=1.45) with time resolution of 12 ms (Andor iXon EMCCD camera). About 1,000,000 images are acquired in each network for statistical analysis. Non-invasive
labeling of chain ends is achieved with hybrid chain construct with a short DNA chain (4 kbp) covalently labeled with dyes grafted onto the end of λ-DNA chains.
CHAPTER 3

HOPS AND STOPS IN DRIVEN TRANSPORT OF POLYMER CHAINS THROUGH HYDROGEL

3.1 Background

Transport in crowded environment is now of emerging interest with examples from granular materials [19–21], colloids [22–24], supercooled liquids [25], and cellular transport [26]. Driven transport of polymer chains threaded through network meshes experience crowded environment as they move around obstacles of network strands. The process has traditionally been modeled as continuous diffusion with directional bias based on ensemble chain mobility [27]. However, empirical observation of individual chains from several simulations and experiments noted intermittent chain motion [28, 29]. Quantitative studies would be highly desirable to understand the transport mechanism. The issues raised are also clearly related to many other diverse non-equilibrium systems with broad implications in separation technology [30], controlled capture and release [31], and delivery through cytoskeleton meshes [32].

3.2 Results

Here we image individual fluorescence-labeled λ-DNA chains pulled by a constant electric force through optically transparent hydrogel network (schematic is shown in Figure 3.1). Although the network is mostly water, its crosslinked network strands presents a crowded environment as a chain (~16 μm in con-
Figure 3.1: DNA chains are negatively-charged at the phosphate groups along the backbone. Agarose gel fibers are not electrically charged. Upon application of an electric force, DNA chains migrate towards the cathode. The chain length) threads through many network meshes (≈200 nm) to be tightly entangled. The DNA chain is uniformly charged along the backbone and the network has no charge so electric force pulls only on the chain. In contrast to classical expectation of continuous transport, we observe intermittent chain motion with alternating hops and stops. The stops show local rattling motion with chain center-of-mass position confined to the mesh size with negligible net displacement whereas the hops transport a chain through the network (Figure 3.2c). Seeking to better understand the transport mechanism, we use a multiscale wavelet method to distinguish hops versus stops for separate analysis from thousands of chains [13]. This allows us to measure kinetics and distribution that was elusive and indirectly inferred in the past. In particular, we show that the dispersion along the main transport direction is greatly enhanced by force and switches from subdiffusive to superdiffusive dynamics with increasing force contrary to classic diffusion assumption. Hopping is an activated process and external force shapes the energy landscape by simultaneously decreasing the overall energy barrier height and skipping a subset of barriers in a heterogeneous environment. Since ensemble-averaged
Figure 3.2: The driven transport of λ-DNA chains through agarose hydrogel network shows intermittent dynamics with alternating hops and stops. a) Average displacement in the field direction against time. Symbols from top to bottom: 16 V/cm, 12 V/cm, 9 V/cm, 6 V/cm, 4 V/cm, and 2 V/cm. (b) A representative time trace of velocity in the field direction $v_x$ (gray curve) showing fluctuations with time at 8 V/cm. $v_x$ averaged over 1 s time window (dark curve) is smooth compared to raw data. The hops (green) and stops (red) are identified with multiscale wavelet analysis. Dashed line denotes the position of zero velocity. (c) The center-of-mass footprints of the same chain as in (b) in $xy$ plane. Color is the same as in (b). (Inset) Magnified view of a region in (c). Scale bar: 0.1 μm.

mobility could be approximately described with continuous transport models, we explore the advantages and limitations of applying classic equilibrium theories to driven systems.

In the long-chain regime where a chain threads through many meshes of the network, the driven motion was thought as continuous process of directed transport [27]. Indeed the ensemble-averaged displacement is linear with time across a wide range of electric field suggesting a constant velocity during transport (Figure 3.2a). However, a closer look at individual chains reveals intermittent dynamics (Figure 3.2b). The instantaneous velocity of a given chain fluctuates a lot with alternating hops and stops. Each hop shows a burst of motion with acceleration and deceleration. Each stop lasts
Figure 3.3: The actual velocity of hops is faster than ensemble velocity. (a) The velocity $v_x$ plotted against $E^2$ where $E$ is field strength in V/cm. Actual velocity of hops (red) is faster than ensemble velocity averaged over both hops and stops (black). Dashed line shows scaling prediction of ensemble velocity $v_x \sim E^2$ for comparison. (b) The ratio of actual velocity to ensemble velocity as a function of $E$. (c) The fraction of time spent in hopping as a function of $E$. Dashed line is guide to the eye.

for several seconds with almost zero velocities although a constant electric force is applied to the chain. The stops despite pulling force and bursts of hopping motion suggest the presence of transport barrier in the network. Classic models of continuous diffusion with direction bias do not describe the intermittent dynamics at single chain level because they are based on ensemble-averaged velocity measurements.

Using a wavelet method, we reliably separate hops versus stops in each single-chain trajectory and rapidly process thousands of chains [13]. The accuracy of separation is confirmed in footprints of a representative chain as shown in Figure 3.2c: the local rattling motion is confined to a length scale comparable to the network mesh size ($\sim 200$ nm) with negligible net
displacement (Figure 3.2c, inset); only hops contribute to transport as they essentially trace out the entire migration path. With efficient separation between hops and stops, we find that the actual velocity of hops is faster than ensemble velocity averaging over both hops and stops (Figure 3.3a), with the largest difference seen at low fields (Figure 3.3b). This is consistent with how field changes the relative amount of time a chain spends between hops and stops. The time fraction of hops increases from a few percent at low field to almost all the time at high field (Figure 3.3c). Therefore, accelerated transport with field arises from a combination of faster hops and fewer stops, in contrast to prediction of ensemble velocity as $v \sim E^2$ suggested by theories of continuous transport. The ensemble velocity we measure also deviates from the predicted scaling (Figure 3.3a), further suggesting the continuous transport description is approximate.

Because chains display fluctuating velocity, we quantify its effect on transport dispersion along the force direction. If the burst-like dynamics reflects thermal fluctuations around ensemble velocity, in other words, a random walk superposed with steady position drift, the displacement dispersion should be diffusive and independent of the directed transport velocity [33–36]. Contrary to that expectation, the displacement dispersion at small force is comparable to equilibrium condition in the absence of force whereas the dispersion at large force is greatly enhanced over several orders of magnitude (Figure 3.4a), further confirming that the observed velocity fluctuation is not trivially due to thermal noise. The force-enhanced dispersion was seen in simulation but modeled as diffusive process with a well-defined dispersion coefficient. In contrast, we observe that dispersion shows unexpected anomalous dynamics (Figure 3.4a). It switches from sub-diffusive at low field to super-diffusive at high field with a limiting scaling exponent of 1.5. We note that similar
exponent was previously observed in simulation of supercooled liquids and attributed to transient trapping [23, 24]. It is reasonable to expect similar mechanism is in effect here. At any instant some chains move forward

Figure 3.4: The dispersion in the field direction shows anomalous dynamics. (a) The dispersion $\sigma^2_x = \langle \Delta x^2 \rangle - \langle \Delta x \rangle^2$ as a function of time. Symbols from top to bottom correspond to 16 V/cm, 12 V/cm, 9 V/cm, 6 V/cm, 4 V/cm, and 2 V/cm. Gray line represents the equilibrium dispersion. Dashed line with a scaling of 1.5 is shown for comparison. (b) Displacement distribution over 0.1 s (gray), 0.6 s (blue), 2.4 s (orange), and 4.8 s (black) at 4 V/cm. Area under each curve is the same. (Inset) Distribution plotted on log scale. (c) Displacement distribution over 0.1 s (gray), 0.6 s (blue), 2.4 s (orange), and 4.8 s (black) at 9 V/cm. Area under each curve is the same. (Inset) Distribution plotted on log scale.
whereas others are stuck transiently; higher field leads to faster transport for the moving subpopulation and hence a larger dispersion. The split of mobility among chains is further confirmed with displacement distribution (Figure 3.4b,c), in sharp contrast to classic assumption of Gaussian displacement distribution for dispersion. At low field a peak near zero displacement represents a transiently stuck subpopulation although a smaller peak gradually shifts in position suggesting a more mobile subpopulation (Figure 3.4b). At high field a shoulder bump develops denoting a slow subpopulation while the main peak shifts as other chains move forward (Figure 3.4c). The distribution plotted on semi-log scale further accentuates the presence of different subpopulations (Figure 3.4b,c, inset). We note that the individual chain variability represents transient dynamic heterogeneity over the observation time windows rather than fixed subpopulations of different mobility as direct imaging of single chains shows that each chain switches repeatedly between hops and stops.

Separation between hops and stops from large data sets allows us to directly measure and quantify the dwell time between hops. Kinetics information of this kind is critical to understand the underlying transport mechanism but could only be inferred in previous experiments from hopping frequency or effective mobility [15, 37]. The distribution of dwell time is exponential for a given field suggesting that hopping motion is an activated process overcoming a barrier with a characteristic rate (Figure 3.5c). The barrier is lowered by force as suggested by the decreasing dwell time with increasing electric field. An earlier study shows that transient stop is not due to chain hooking on network strands [38]. Here we find dwell time quantitatively consistent with entropic trapping model (Figure 3.5b) which assumes a chain overcome entropic barrier to squeeze through narrow constriction according
to the random crosslinking between network strands. In this model, due to the competition between entropy cost and electric energy gain of pulling a chain along the field though the network, the height of the energy barrier is inversely proportional to field and therefore shorter dwell time with increasing field is predicted as $\ln \tau \sim 1/E$ with a frequency factor of $\sim 0.2$ s and barrier height of $\sim 8.4/E \ kT$ at each field [37, 39]. The entropy description is unexpected for a whole chain threading through many meshes because entropic trapping is typically considered in a regime where chain size is comparable to mesh size [40]. Earlier studies show that driven motion is initiated at a chain end which stretches out after being localized for some time in the

Figure 3.5: The dwell time between hops shows a field-activated transport mechanism. (a) Dwell time between hops decreases with field. Dashed line is guide to the eye. (b) Same data as (a) plotted to show $\ln \tau \sim 1/E$. Dashed line is a linear fit for comparison. (c) Dwell time is exponentially distributed. Examples from 4 V/cm, 6 V/cm, and 9 V/cm are shown.
network [38]. We therefore consider that the effective entropic barrier to a migrating chain might be imposed locally at the leading chain end and intra-chain tension facilitates the rest of the chain to follow with less resistance to motion. Alternative to entropic squeezing, if we consider DNA chain as worm-like-chain with persistence length on the order of mesh size as in this system, it might have trouble wriggling through rigid network meshes. It is expected that chains deform more easily with larger force and can wriggle through meshes, although we cannot find studies to estimate energy barrier considering the chain stiffness in the confinement of network meshes.

Plotting average run length between stops, we find the run length increases with field drastically (Figure 3.6a). The run duration also increases with field accordingly (Figure 3.6b). As stopping location maps the location of the effective barriers, the increasing distance between stops suggests the effective barriers become sparser with field.

3.2.1 On energy landscape

3.3 Discussion

It is likely that a minimum barrier height is required to hinder transport while the energy landscape has a distribution of barrier height given the heterogeneous local hydrogel environment. As electric field lowers the barriers, some barriers that hinder transport at low field would become below the threshold at high field. Based on the distance between effective barriers, a threshold barrier of \( \sim 2 kT \) is consistently found assuming various forms of barrier height distributions (Figure 3.7). Compared to typical microfluidics with features that have strictly a single barrier height [37], the transport
through network has a stronger field dependence through both decreasing dwell time and skipping a subset of barriers. Indeed the steep rise with field for fraction of time spent in transport phase cannot be explained with decreasing dwell time alone (Figure 3.8). We note the distribution of barrier height might have a subtle effect on the hopping kinetics as it can still be
Figure 3.7: Different energy barrier distribution gives consistent threshold barrier of $\sim 2\ kT$. Data is the fraction of effective barriers compared to total number of barriers. Red line denotes exponential distribution of barrier height which estimates a threshold barrier of $\sim 1.7\ kT$. Yellow line denotes Gaussian distribution with a width of $1\ kT$ which estimates a threshold barrier of $\sim 2.3\ kT$. Blue line denotes Gaussian distribution with a width equal to average barrier height which estimates a threshold barrier of $\sim 2.2\ kT$. The fraction is calculated based on the distance between detected barriers and assuming the barrier size as mesh size of $\sim 0.2\ \mu m$.

described with an effective single barrier (Figure 3.5c).
Figure 3.8: Decreasing of waiting time alone cannot explain the steep increase of moving fraction. Red line is the predicted moving fraction with the effect of decreasing waiting time.

3.3.1 Connection to other driven systems

It is recently shown burst-like dynamics often has chaotic behavior in various viscoelastic systems [21, 25, 26]. In contrast, although we observe similar intermittent motion with bursts of hopping, the transport here is a stochastic process as run length, run duration, and dwell time are all exponentially distributed (Figure 3.6, inset; Figure 3.5c); the run displacement is linear with run time a well-defined run speed (Figure 3.3). The unique kinetics is likely due to static hydrogel network that does not relax and rearrange in
response to driven transport, in contrast to other viscoelastic media.

3.3.2 Wavelet parameter choice to identify the two states

While to the eye it may be obvious that molecular mobility exhibits two states (Figure 3.9a), to automatically separate these two is challenging, as the frame-to-frame displacements of center of mass shows no temporal pattern above random fluctuations (Figure 3.9b). To discriminate these two mobility states, a wavelet analysis was used on each single-molecule trajectory. The detailed description of the wavelet can be found elsewhere [13]. An advantage of the wavelet method is that it automatically determines the noise level based on the raw data on short time scale without setting an arbitrary threshold. One can adjust the scale (the time window to examine the average drift in position) to optimize the detection. We show that the detection of the two states is robust across a wide range of parameter space, as shown in Figure 3.9. On scales 8 and 32, the universal threshold separated the hops and stops nicely for \( \lambda \)-DNA in 1.5 wt\% agarose gel under electric field 12 and 6 V/cm respectively (Figure 3.9a and c). At each field strength, a wide range of scales give good separation (Figure 3.9c). Too-small scales result in missing a large portion of spurts, too-large scales assign mistakenly many stops as spurts. The envelope of usable scales (which still spans a broad range) decreased with field strength because, as the lifetime of pausing shortened as the force on the DNA molecules increased, shorter scale became better at most accurately assigning these faster transitions. The speed during spurts much exceeded the mean speed (Figure 3.9a). This significant contrast would not have been quantified otherwise, and provides firm numbers from which to examine competing electrophoresis theories.
Figure 3.9: Detection of the two states is robust across a wide range of parameter space. (a) Illustrative trajectories showing that DNA center of mass motion at 12 V/cm is discontinuous, the wavelet analysis identifying spurts of rapid motion (red) and pauses between spurts (gray), neither of them equal to the mean speed. The trajectory at 6 V/cm drive is discontinuous likewise. (b) Frame-to-frame displacement of a 6 V/cm trajectory (33 fps) is plotted against time. (c) This panel compares efficacy over a broad range of scale of analysis as well as drive voltage. Over a broad intermediate range of scale the mobility separation of this electrophoresis data is robust without depending on the specific choice of scale.

3.4 Conclusions

Imaging individual fluorescence-labeled λ-DNA chains (∼16 µm in length) threading through agarose hydrogel network meshes (98.5% water content
with average mesh size \( \sim 200 \text{ nm} \) when pulled by a constant electric force, we observe that chain motion is intermittent with alternating hops and stops. The stops show local rattling motion with center-of-mass position confined to the mesh size whereas the hops show burst-like dynamics which lead to directed transport through the network. Separate analysis on hops versus stops from thousands of chains provides direct kinetics and distribution information that was previously inaccessible. We show that hops are activated process overcoming transport barrier with a characteristic force-dependent rate, with shorter dwell time between hops with larger force. The effective barriers become sparser with larger force suggesting that force shapes the energy landscape by simultaneously decreasing the overall energy barrier height and skipping a subset of barriers in a heterogeneous environment. The dispersion of displacement along force direction is greatly enhanced by force and switches from subdiffusion at low field to superdiffusion at high field, contrary to classic diffusion assumption.

3.5 Materials and Methods

3.5.1 Fluorescence microscopy

Data were acquired in epifluorescence mode, typically at a frame rate of 33 fps. A 532 nm excitation laser was focused at the rear focal point of an oil immersion objective (Zeiss, \( \alpha \)-Plan Fluor 100\( \times \), NA = 1.45) with 2.5\( \times \) post-magnification to image with a resolution of 64 nm \( \times \) 64 nm per pixel. Fluorescence images were collected through the same objective and detected by a back-illuminated electron multiplying charge-coupled device (EMCCD) camera (Andor iXon DV-897 BV) after filtering out light from the excitation
laser. The movies were converted into digital format and analyzed. A typical dataset consists of 30 movies at each field condition, each of them consisting of 4,000 frames per movie acquired at 33 fps. The resulting dataset of conformations typically amounts to $>10^4$ from thousands of molecules.

3.5.2 DNA samples

$\lambda$-DNA (48.5 kbp, Promega) was labeled by covalently attaching dye, a RhB derivative (Mirus) to heteroatoms on DNA, at a labeling density of roughly one dye per 5 base pairs. Single-molecule measurements of DNA chain conformations were made in a miniature gel electrophoresis setup using agarose gel (final concentration 1.5% (w/v)), in 0.5× TBE buffer (45 mM Tris, 45 mM borate, 1 mM EDTA), the DNA being at picomolar concentration. Anti-photobleaching agent, ascorbic acid (SigmaAldrich), was present at a final concentration of 10 mM. A DC voltage was applied across two Pt electrodes to generate an electric field ranging from 6 to 16 V/cm.

3.5.3 Miniature gel electrophoresis cell for imaging

A miniature gel electrophoresis setup was constructed to allow direct imaging in epi-fluorescence mode. Test chains were added to agarose suspension at 45 °C at picomolar concentration for single molecule imaging. 1 ml solution was cast into an imaging sample chamber (Lab-Tek, bottom area 20 mm×20 mm) at 45 °C and allowed to solidify at room temperature for 20 min. Both sides of the gel pad were cut and removed to accommodate two Au electrodes running parallel along the gel sides. The Au electrodes were made by electron beam deposition of a 200 nm Au layer onto the glass substrate. The solidified agarose gel slab was immersed in 3 ml buffer in a so-called
“submarine” mode. Electrophoresis buffer (1× TBE) was present to ensure
electrical circulation and prevent electrolysis. A mixture of glucose, glucose
oxidase, calatase and trolox was used to minimize photobleaching. The sample
chamber was mounted onto a microscope stage. A DC voltage was applied
across the two electrodes through a function generator (Wavetek) to generate
an electric field up to 16 V/cm.
CHAPTER 4

ZIG-ZAG PATHS WHEN DNA SPEEDS THROUGH HYDROGEL

4.1 Background

Dispersion is ubiquitous: from mundane color of cloud and milk to sophisticated technologies such as ultrasound imaging and scanning electron microscope. The process of dispersion happens when directed transport is obstructed and an object turns to the transverse direction, perpendicular to transport, to move around obstacles. This transverse motion is also commonly seen in many microscopic systems (e.g. bacteria chemotaxis or moving through porous media [41–44]; intruder driven through granular materials [19, 20, 22]; cellular transport of vesicles [26]). Regardless of the origin, transverse dispersion has important implications in many diverse physical systems such as mixing and de-mixing streams of fluids in reactors [45], efficient separation in chromatography [46, 47], and transport through cellular cytoskeleton meshes or other crowded environment [25, 30, 32]. The vast amount of past literature studies the ensemble-averaged behavior from multiple dispersion events such as intensity profiles [48, 49], penetration depth [50, 51], and temporal fluctuations of intensity [52, 53]. However, direct observation of individual dispersion events that give rise to the ensemble-averaged behavior has been lacking.

We image hundreds of single λ-DNA chains driven through agarose hydrogel network by an electric field where the footprint of a chain follows a
tortuous path, effectively described as successive individual dispersion events at defined turning points. The network acts as fixed obstacles and forces a chain to go around and in between strands. Compared to other transport systems that exhibit transverse motion, the network in this system does not relax or rearrange to complicate interpretation of chain dynamics. The voluminous literature has extensively studied ensemble-averaged chain transport through network along the external force direction, but too little is known from experiments about chain motion perpendicular to the force due to the paucity of high-throughput imaging of single chains.

4.2 Results

λ-DNA chains are covalently labeled with fluorescence dyes uniformly along the chain and tracked as they are driven through optically transparent agarose network by a DC electric field. On average a chain threads through dozens of meshes inside the network. Previous studies show that the whole chain follows the path of its leading end [38], we therefore track the center-of-mass motion to study chain transport. The footprint a migrating chain lays down is a tortuous path although a constant electric force is applied to the chain (gray tracks in Figure 4.1a,b). The overall chain motion aligns with force direction but it turns back and forth in the direction perpendicular to force on length scale a few times the average mesh size (∼200 nm). To quantify the orientation bias and its dynamics, we first identify turning points along the track where large direction change occurs, denoted with red dots in Figure 4.1a,b. Overlaying with the original track confirms that the turning points follow the evolution of chain orientation. In other words, the motion can be described effectively as successive individual dispersion events.
at defined turning points that deflect the otherwise straight and directed driven motion.

Interestingly, the transverse motion is distinct between high and low field. At low field the chain motion is often at a notable angle with the force and shows frequent local motion along the track and occasional backward motion against the force direction. In contrast, at high field the motion is more directed with a straighter and smoother track showing less local or backward motion. It suggests that some paths in the network inaccessible to a chain at low field become available at high field as higher field might allow a chain to channel through narrow constrictions in the network more easily although the spatial configuration of static network obstacles is unchanged [9, 10]. The angle between motion and field direction is calculated based

Figure 4.1: a,b) Tortuous paths (gray) of DNA chains driven by electric field through network can be effectively described as individual dispersion events at defined turning points (red) along the track. Higher field leads to straighter and less tortuous path. Force direction is in x direction and transverse dispersion is in y direction. Angle between motion and force direction is denoted as $\theta$. c) and d) Histogram of angle $\theta$ distribution plotted on a circular plot. e) Mean $\cos \theta$ changes between 0 and 1 with increasing electric field. f) Mean squared angle displacements as a function of time. A slope of 1 representing diffusive dynamics is shown for comparison.
on the turning points and its distribution is plotted as a histogram on a circular plot (Figure 4.1c and Figure 4.1d). Consistent with the transverse dispersion, the angle at low field is more spread out with respect to the field direction with a wide distribution across the first and forth quadrants and occasional backward motion as represented in the second and third quadrants. In comparison, the angle at high field is narrowly distributed in the small angle region in the first and forth quadrants and displays almost no backward angles. By adjusting the field strength, the average orientation can be tuned over a wide range as $\cos \theta$ increases monotonically with field strength between 0 (isotropic) and 1 (perfect alignment) (Figure 4.1e). The mean squared angle displacement (Figure 4.1f) is initially diffusive, similar to dynamics of a generic rotator doing rotational random walk, and gradually plateaus at long time, consistent with the orientation bias exerted by field. The difference between high and low field lies in the bias level, the higher field leading to stronger orientation bias and lower plateau level. A curious feature is that no discernible difference is observed in the initial diffusive regime between high and low field which might suggest a base level of angle fluctuation due to thermal Brownian motion.

Since the turning points identify individual dispersion events, we first measure the transverse displacement perpendicular to the field between individual dispersion events and find that it is exponentially distributed with a characteristic length (Figure 4.2a). Interestingly, the distribution between high and low field overlaps and the characteristic length is independent with field (Figure 4.2a, inset). This length scale is comparable to the network mesh size, suggesting that the network structure might determine the location of chain dispersion and the magnitude of the transverse fluctuations. A reasonable expectation is that dispersion would be a stochastic process that the chain is
Figure 4.2: a) Distribution of transverse displacement $|\Delta y|$ between individual dispersion events. The dashed line is guide to the eye. (Inset) Mean transverse displacement $|\Delta y|$ in $\mu$m as a function of electric field $E$ in V/cm measured based on turning points (■) or direction switching (●). b) Distribution of waiting time between individual dispersion events. (Inset) Mean frequency of scattering in $s^{-1}$ as a function of electric field $E$ in V/cm. Symbols are the same as in (a).
scattered with a characteristic rate constant, similar to many other random walk processes such as run-and-tumble of bacteria with a turning rate and a generic 1D random walker that switches direction at a certain rate. Indeed we find the waiting time between individual dispersion events to be exponential, confirming the stochastic nature of chain dispersion (Figure 4.2b). Interestingly, the scattering frequency becomes faster with increasing field. Since the transport mobility of chain is known to accelerate with field [27], the increased scattering frequency likely reflects faster encounter with obstructing network strands. Therefore, electric field acts like an effective elevated temperature to promote dispersion by increasing the scattering frequency. As a consistency check, we also analyze transverse direction switching, identified as a subset of turning points that switch between $+y$ and $-y$ direction. As expected, similar conclusions are reached that the exponential distribution of distance and waiting time between individual events and general trend with field are preserved (Figure 4.2, insets). The force-enhanced dispersion normal to the force is evident when we plot variance of transverse displacement against time for different field conditions calculated from hundreds of chains (Figure 4.3a). For a given time interval the variance at low field is comparable to without field and shows considerably enhanced dispersion at high field. The acceleration of transverse dispersion by field is further confirmed with a wider distribution of transverse displacement at higher field for a given time interval (Figure 4.3a, inset). The transverse dispersion shows orders of magnitude increase in dynamics with field and does not saturate, a much stronger effect than theoretical predictions [36].

The dynamics of transverse spreading show an approximately linear increase of variance with time (Figure 4.3a). However, a closer inspection of the data when plotted on log-log scale shows anomalous dynamics that
persist several orders of magnitude in time (Figure 4.3b). Instead of strictly Fickian diffusion, the variance is slightly sub-diffusive (scaling exponent

Figure 4.3: a) Variance \( \sigma_y^2 \) of transverse motion against time. The variance in the absence of electric field is shown as a line (purple). Symbols from bottom to top correspond to 2 V/cm, 4 V/cm, 6 V/cm, 9 V/cm, 12 V/cm, and 16 V/cm. (Inset) The distribution of transverse displacement \( \Delta y \) over 1.5 s for 2 V/cm (inner curve) and 16 V/cm (outer curve). Unit of \( \Delta y \) is \( \mu \)m. The distributions are normalized to the maximum. b) Same data as in (a) plotted on log-log scale. Slope of 1 is shown for comparison.
<1) at low field and super-diffusive (scaling exponent >1) at high field. It is known that diffusive dynamics of transverse motion might be expected from stochastic multiple dispersion events as diffusion can arise from directed motion with stochastic rotation to change direction [54] and is observed in simulations of transverse motion of driven particles in granular materials [23, 24]. In fact, simulated 1D random walk trajectories with constant speed and exponentially distributed time between direction switching events matching experimental values can reproduce the observed diffusive spreading dynamics. The discrepancy with observed anomalous dynamics suggests that the dispersion involves more complex processes than previously thought and the dynamics of the transverse motion might be coupled to the driven motion in the orthogonal direction.

The large datasets allow us to scrutinize the distribution of transverse displacement over 3 decades in probability density from hundreds of chains (Figure 4.4a). The distribution spreads out with increasing time interval at a given field, consistent with the time-dependent variance (Figure 4.3). The distributions at different time intervals can approximately collapse, with scaled time. For example, the displacement distributions at 6 V/cm can overlap with $t^{0.5}$ (Figure 4.4b), suggesting the transverse motion is diffusive with the same form of displacement distribution under this condition [55–57]. In fact, the displacement distribution across a wide range of field can collapse with the proper scaling of time, ranging from $t^{0.25}$ to $t^{0.6}$ (Figure 4.4c,d), consistent with anomalous dynamics of sub-diffusive motion at low field and super-diffusive motion at high field. The shape of the distributions significantly deviates from classical Gaussian expectation and shows a heavy tail indicating a large fraction of subpopulation with large transverse displacements (Figure 4.4b-d). The heavy tail suggests the dynamic heterogeneity
of the system reminiscent of supercooled and glassy systems although the underlying physical origin is different. Here, instead of collective motion of a cluster of mobile particles [58–60], the fast subpopulation suggests that the network environment is heterogeneous, even on length scale an order of magnitude larger than the mesh size, with regions containing effective channels that favors transverse chain motion with respect to the force direction. The

Figure 4.4: a) The distribution of transverse displacement ($\Delta y$) spreads out with time. The time interval is 0.1 s ($\Box$), 0.3 s ($\bigcirc$), 0.6 s ($\Delta$), 1.2 s ($\triangledown$), 2.4 s ($\times$), and 3.6 s ($\Diamond$). The distributions are normalized to the maximum and plotted on semi-log scale. b) The distributions of $|\Delta y|$ at different time interval can collapse with normalized displacement $\Delta y/\Delta t^{0.5}$ for 6 V/cm. A Gaussian fitting is shown for comparison (dashed line). Symbols are the same as in (a). c) The distributions of $\Delta y$ at different time interval can collapse with normalized displacement $\Delta y/\Delta t^{0.25}$ for 2 V/cm. A Gaussian fitting is shown for comparison (dashed line). Symbols are the same as in (a). d) The distributions of $\Delta y$ at different time interval can collapse with normalized displacement $\Delta y/\Delta t^{0.6}$ for 16 V/cm. A Gaussian fitting is shown for comparison (dashed line). Symbols are the same as in (a).
non-Gaussian distribution is recently observed in many diverse soft matter systems and its origin is speculated to arise from insufficient averaging of spatial and temporal heterogeneity [55–57]. We find that the distribution is closer to Gaussian at higher field suggesting that a chain experiences more collisions to randomize the transverse motion or the channeling effect of a subpopulation of chains is less important at high field.

4.3 Discussion

Dispersion in the direction normal to directed transport is common but direct observation of the dispersion process at single-molecule level is lacking. Here, λ-DNA chains are labeled with fluorescence dyes and directly imaged with fluorescence microscopy with millisecond and nanometer resolution as chains are driven through agarose network by a DC electric field. The footprints of a chain follow a tortuous path exhibiting prominent transverse dispersion perpendicular to the force. Analyzing hundreds of chains, we find that the field acts like an effective elevated temperature to speed up transverse dispersion by increasing microscopic scattering frequency. Contrary to classic diffusion assumption, the dynamics of dispersion is anomalous with sub-diffusion at low field and super-diffusion at high field. The distribution of transverse displacement evaluated over 3 orders of magnitude can be collapsed with proper normalization of time between $t^{0.25}$ and $t^{0.6}$ across a wide range of field. The heavy-tailed distribution suggests unexpected dynamic heterogeneity with much more frequent large displacements, suggesting a underlying network with complex local structure. The study points to possible routes to actively control the dynamics of transverse dispersion with external forces. Analyzing thousands of individual dispersion events with millisecond and nanometer
resolution, we find that external force acts like an effective temperature to speed up the transverse fluctuation by increasing the microscopic scattering frequency. The distribution of transverse displacement over three orders of magnitude shows large displacements much more frequent than expected. The results raise questions concerning classical assumptions and modeling of transverse dispersion in driven systems.

Figure 4.5: Simulate 1D random walk trajectories with constant speed and exponentially distributed time between direction switching events matching experimental values can reproduce the observed diffusive spreading dynamics. Symbols are from experiments and lines are from simulations. The lines from top to bottom corresponds to field conditions at 16 V/cm, 12 V/cm, 6 V/cm, and 3 V/cm.
4.3.1 Connection to 1D random walk

It is known that diffusive dynamics of transverse motion might be expected from stochastic multiple dispersion events as diffusion can arise from directed motion with stochastic rotation to change direction and is observed in simulations of transverse motion of driven particles in granular materials. We therefore simulate 1D random walk trajectories with constant speed and exponentially distributed time between direction switching events matching experimental values and find that they can reproduce the observed diffusive spreading dynamics, as shown in Figure 4.5.

4.3.2 Connection to other driven systems

Looking to the future, it remains to be tested whether the dispersion features revealed in the current study are relevant to other dispersion systems, for example, when in a dynamic environment such as granular materials or fluid systems, or when chains are replaced by strictly rigid body to remove internal conformation fluctuations, or when the microenvironment is ordered such as fabricated nanopost arrays [61]. To this end, we attempted to embed charged fluorescent particles into hydrogel networks. However, there are several technical challenges that need to be overcome. First, the surface chemistry of particles could be tricky and special effort is needed to ensure no specific adhesion attraction between particles and the hydrogel network. From past experience, particles fully covered with poly-ethylene glycol (PEG) chains are most suitable for this purpose. Second, the particles should be highly charged to be driven through the hydrogel with directed motion. The various PEG particles we used were too weakly charged for that purpose.

The field enhancing effect suggests a possible route to actively control the
dispersion dynamics normal to the transport direction, relevant to processes such as mixing, separation, and transport through crowded environment.

4.4 Conclusions

Single-molecule imaging shows zig-zag paths when electric fields drive DNA up to orders of magnitude faster than Brownian diffusion through hydrogel with which is it entangled (contour length 90 times longer than the average network mesh size of 200 nm) causing molecules to appear to diffuse normal to the field direction. The higher the driving field, the higher this transverse dispersion, shifting from subdiffusive to superdiffusive with increasing field. In analogy to common Brownian diffusion at equilibrium, we measure in this driven situation the so-called van Hove distribution in the off-axis direction. This displacement distribution, we quantify over 3 orders of magnitude of probability density, finding master curves that normalize with elapsed time. They likely reflect how moving DNA scatters off the disordered static gel through which it passes, in rough analogy to how light passes through smoke and other instances of multiple scattering.

4.5 Materials and Methods

4.5.1 Fluorescence microscopy

Data were acquired in epifluorescence mode, typically at a frame rate of 33 fps. A 532 nm excitation laser was focused at the rear focal point of an oil immersion objective (Zeiss, α-Plan Fluar 100×, NA = 1.45) with 2.5× post-magnification to image with a resolution of 64 nm × 64 nm per pixel. Fluorescence images were collected through the same objective and detected
by a back-illuminated electron multiplying charge-coupled device (EMCCD) camera (Andor iXon DV-897 BV) after filtering out light from the excitation laser. The movies were converted into digital format and analyzed. A typical dataset consists of 30 movies at one field condition, each of them consisting of 4,000 frames per movie acquired at 33 fps. The resulting dataset of conformations typically amounts to $>10^4$ from thousands of molecules.

4.5.2 Sample preparation for imaging

λ-DNA (48.5 kbp, Promega) was labeled by covalently attaching dye, a RhB derivative (Mirus) to heteroatoms on DNA, at a labeling density of roughly one dye per 5 base pairs. Single-molecule measurements of DNA chain conformations were made in a miniature gel electrophoresis setup using agarose gel (final concentration 1.5% (w/v)), in 0.5× TBE buffer (45 mM Tris, 45 mM borate, 1 mM EDTA), the DNA being at picomolar concentration. Anti-photobleaching agent, ascorbic acid (SigmaAldrich), was present at a final concentration of 10 mM. A DC voltage was applied across two Pt electrodes to generate an electric field ranging from 6 to 16 V/cm.

4.5.3 Algorithm to identify turning points on single-molecule trajectories

As the path a migrating chain follows inside the gel is tortuous although a constant force is applied to the chain, we identify turning points to coarse-grain trajectories to remove noise from the raw data while keeping the orientation information as we did in Figure 4.1. The analysis of individual scattering events is based on identification of turning points. To define turn-to-turn steps, we used the error radius method documented in the literature
In this method, a straight line connects two points along the trajectory, then one end of it extends to subsequent data points in order until the data points in between start to deviate from the line above a threshold. The last end point is determined as a turning point, and is used to start a new line. The threshold for each trajectory is adapted to the end-to-end distance of that particular trajectory, to account for the heterogeneity between trajectories. The threshold converges when the turning angles become uncorrelated and isotropic distributed.
5.1 Background

This chapter focuses on methods of image analysis that enable one to go beyond tracking the position of single molecules, to also analyze their internal conformations, in instances when molecules display shape fluctuations under various conditions ranging from thermal equilibrium to deformation under mechanical stress or other external field. As these changes are rapid, we are interested in real-time measurements during which the need to acquire data rapidly without signal averaging introduces experimental uncertainty. These problems of tracking internal degrees of freedom, which become technically feasible when the size of macromolecules exceeds the resolution of a microscope in one or more directions, apply especially to tracking biological macromolecules, among them filaments such as actin [11, 12] and more flexible molecules such as DNA [63–71].

Methods of image analysis could not be applied until recently to problems of this kind; indeed, in the early days the data were typically acquired by video microscopy, as for example in measurements of DNA and actin using fluorescence microscopy [11, 12, 63–72]. Almost from the beginning, special

---

This chapter is partially reproduced from Guan, J; Wang, B; Granick, S. Automated single-molecule imaging to track DNA shape. *Langmuir* **2011**, *27*, 6149-6154.
attention was given to direct observation of polymer conformations perturbed from equilibrium by mechanical force or electric field, but quantification was held back in part by the limited resolution of video cameras, in part by the inability of routinely analyze images using methods that would require significant computing power and data storage capacity. It is understandable that quantification to date has concerned largely radius of gyration and coarse measurement of shape anisotropy by measures such as the long axis and short axis components of fluorescence images in two-dimensional projection in the plane of a microscope [2, 17].

In addition, many of the early studies suffer from few statistics and involve analyzing a small number of images [70–72]. Yet from the beginning of this line of research, it was evident [63–70] that single-molecule analysis of chain conformations holds the promise to measure the distribution of chain conformations whose averages enter into important ensemble-averaged quantities, such as rheology and electrophoretic mobility. Nowadays, it is feasible to use inexpensive personal computers to facilitate analysis of polymer conformations with large statistics and high accuracy. This study is in the spirit of an earlier pioneering automated line tracking method, introduced to analyze actin [73], that works well to analyze filamentous macromolecules that are stiff. The image analysis methods described below are designed to apply equally to more flexible macromolecules such as DNA.

The focus here is primarily on the methods of automated image analysis we have developed to deal with this problem and to explain the logic that prompted the choice of those methods rather than others. These methods consist of three elementary stages: first, to discriminate the shapes of macromolecules from noise, which we refer to as “feature finding”; second, to approximate those shapes as lines, which we refer to as “line tracking”;
third, to discriminate reasonable from unreasonable fitted conformations in
the time domain, which we refer to as “temporal consistency check.” Finally,
this paper presents two examples of applying the method. Although the
methods are general and should readily be adaptable to other visualization
techniques, here we illustrate the arguments using sample images from DNA
in a fluorescent microscope. We focus on DNA that is driven by electric fields
to adopt relatively extended conformations and a variety of shapes.

These methods, which build upon image processing methods reported ear-
lier from this laboratory which at that time did not consider internal degrees
of freedom in the tracking of single molecules [74, 75], amount to approxi-
mating the true shape of linear macromolecules by renormalized curved lines,
lines that are optimized to describe the shape within the limits of optical res-
olution. As methods of super-resolution imaging [76] gain more widespread
use, one can expect even more need for image analysis methods to analyze
internal chain conformations.

5.2 Procedures of line tracking

The purpose here is to quantify, within the limits of optical resolution, the
linear shapes of CCD images that are noisy, faint, and diffraction-blurred. In
the methods described below, we take the approach that it is better to reject
data from consideration than to improperly include it in subsequent analysis.
Therefore, especially in the third step of analysis consisting of checks of data
consistency in the time domain, we reject up to 90% of the data acquisition
frames. Examples of the variety of raw data are given in Figure 5.1, panels
c, d, and e.
5.2.1 Identifying each DNA molecule

For various practical reasons, despite the most careful efforts to optimize image quality, the signal-to-noise ratio is never satisfactory in unprocessed images (Figure 5.1a). First, the finite DNA concentration results in background fluorescence from other molecules that contribute to the image in spite of being located outside the focal plane. Secondly, the DNA that we track moves rapidly under electrophoresis, leaving limited time to collect

Figure 5.1: Examples of raw data in which λ-DNA (contour length 16 µm) displays various conformations in an optical microscope. Color bar corresponds to relative intensity within each image. (a) A typical unprocessed image as input raw data. (b) A local Gaussian filter reduces noise significantly. (c),(d),(e) Additional examples of various polymer conformations with final line tracking results (black line) overlaid. Scale bar: 1 µm.
photons at each position. Third, the elongation of chains under electric field lessens the local fluorophore density. We find that applying a Gaussian filter with a width of 1 pixel locally at each pixel, with Gaussian weighted contribution of intensity from neighboring pixels, significantly reduces noise without compromising the main features of a DNA molecule (Figure 5.1b).

As each image consists primarily of background, only a few pixels representing the dilute DNA molecules of interest, we estimate background intensity noise from the mean intensity at each pixel, and estimate noise level from the variance of these pixel intensities. After subtracting this background from each pixel in the image, we retain for subsequent analysis only pixels at which the residual intensity exceeds a threshold, typically set to be 710 times the noise level. Control checks show that subsequent line tracking does not depend sensitively on the detailed choice of threshold value, nor on whether the image of the DNA is included in estimation of the background.

Next, the image analysis makes judgments to decide whether two features close in space belong to a single molecule. It is a problematical question because low signal-to-noise ratio pixels may in principle register the absence of DNA, but also in principle may indicate that low intensity parts of the molecule, in some cases even vanishingly-low intensity due to various reasons discussed earlier in this section. Figure 5.2a illustrates a stretched DNA whose middle part presented an intensity less than an intensity threshold. To analyze this situation, first we connect pixels whose distance is less than a threshold and consider tentatively the connected pixels to come from a single molecule. Depending on the threshold selected, the pixels in an image then cluster into either a single grouping, or several. For example, using a threshold distance of 8 pixels, Figure 5.2b presents two clusters of pixels, whereas a threshold of 9 pixels implies just a single cluster of pixels; the
difference depends on whether critical connections are allowed or disallowed by that threshold (white lines in Figure 5.2b). To automate the process of line tracking, in order to perform this grouping of pixels first we apply a fixed threshold and analyze the best line through the implied grouping. Then unreasonable lines are removed at a later stage of analysis through the automated temporal consistency checks discussed in Section 5.2.2.

Figure 5.2: Feature finding: grouping pixels into one DNA molecule or several according to the threshold distance between pixels. (a) An example of an image in which only a fraction of the middle pixels have high enough intensity to be recognized as signals. (b) The connections between pixels discriminate whether they cluster into one grouping or several. The leftmost two white connections are critical if not present, this image is treated as two groupings, each one from a different molecule. The pink connections are separated by 5-8 pixels. The white connections are separated by 8-9 pixels. Scale bar: 1 µm. Color bar is the same as in Figure 5.1.
5.2.2 Tracing a line through each DNA molecule

This relies on the concept of a minimum spanning tree, a concept in graph theory which quantifies the shortest path length between nodes in an image [77]. Here, the nodes in the image are the pixels of the molecule defined from the previous section. Figure 5.3a illustrates the concept schematically and Figure 5.3b illustrates it for a given set of data acquired in our experiments. First, a tree is constructed by connecting every two pixels. However, since some pixels are bright and others are dim, and we wish to preferentially include pixels of high intensity to minimize the chance of resulting lines being trapped on noise pixels, the connection length between pixels is assigned not simply as the spatial distance by which they are separated, but weighted by the inverse of their sum intensity. A minimum spanning tree is then generated choosing from existing connections based on each connection length. We find that this feature of intensity weighing is often necessary but that the exact intensity weighing method is not critical in line tracking; for example, an alternative weighting, exponential weighting according to intensity, gives similar results. Searching through this minimum spanning tree, one can find one path between two termini that contains the largest number of pixels: this is the longest path through this molecule (Figure 5.3c).

To yield the final line tracking result (Figure 5.4a), we follow a four-step progressive refinement of the result. First, to smoothen the line, we perform a polynomial fit, typically a quadratic fit, locally on adjacent pixels (Figure 5.4b). The local fit is preferable to fitting the entire line with a polynomial because the overall shape is sometimes highly curved. An additional advantage of local fitting is that to improve accuracy, one can select at each point which coordinate, horizontal x or vertical y, to fix or fit. The choice to
fix or fit is done by performing both x and y fits and choosing automatically one fit to minimize local fitting error. After this automated process, each pixel is reassigned a new point that has a modified x or y position. We think this local fitting with flexible x or y is more faithful to the original shape than a single global polynomial fit and sometimes the only possible way to fit a highly curved shape.

To further smoothen the line, the second step is to fit the cross-section

Figure 5.3: Line tracking: using minimum spanning tree analysis to find a longest path through pixels to identify a line through the DNA molecule. (a) Schematic illustration of the notion of minimum spanning tree. The edges belonging to the minimum spanning tree are drawn in solid lines whereas other edges are drawn with dashed lines. Next to each edge, the indicated numbers specify the relative weight. (b) Example of an intensity-weighted minimum spanning tree (white) overlaid onto real data. (c) For the data in panel (b), a longest path (thick white line) is identified from analyzing the minimum spanning tree. Scale bar: 1 µm.
intensity profile at each point along it to a Gaussian intensity profile, the
principle being that diffraction-limited images are expected to be described
by this function. The position of each pixel along the line is then accord-
ingly adjusted to the center of the Gaussian (Figure 5.4c); typically, this
adjustment is on the order of one pixel. In this fitting, the orientation of the
cross section is taken to be perpendicular to the tangent line at each point.
As the coordinates at this stage of the analysis are typically not located on
integer pixel positions, the intensity at each point is calculated based on the
surrounding four pixels (integer coordinates) via linear interpolation.

Third, the minimum spanning tree is recalculated from the points which
resulted from the second step (Figure 5.4d). The process to do so is the same
as before, except that adjacent points are linked without intensity weighing.
This step is to deal with occasional failures in the second step to find an
accurate tangent direction and hence the correct cross section. For example,
the radical jag in Figure 5.4c came from this occasional failure. Though the
line looks jagged in this zoom-in presentation, this new position deviates less
than two pixels from the original. We exclude these anomalous positions
nonetheless using a minimum spanning tree.

Last, high-frequency noise is removed from the line using a wavelet filter
(Figure 5.4e). It is convenient to employ a discrete wavelet transform based
on Daubechies-16 wavelet of the line. The level 2 coefficients, which are
the decomposition in high frequency range corresponding to neighboring 2-
4 points, are soft thresholded with coefficients exceeding the threshold, δ,
in magnitude were pushed toward zero by δ, while those less than or equal
to were set to zero. The threshold is calculated through universal threshold,
δ = ̂σ√2lnN, where N is total number of points and ̂σ is estimated via median
absolute deviation (MAD), ̂σ ≡ median(|x_i|_{i=1,...,N})/0.6745. Level 2 is chosen since

67
noise is expected to dominate the high frequency. The choice of coefficients, working at the level of neighboring 2 to 4 points, is primarily to remove high frequency noise in the line. The smoothened line is obtained by inverse transforming the coefficients after thresholding.

Figure 5.4: Four-step refinement of the line identified in each single image. The highlighted region in panel a, magnified in panels b-e, is used to illustrate how a line changes from each step to the next step during the refinement procedure. In panels b-e, starting lines and symbols are white, and the line resulting from that step is drawn in black. (a) Starting from a longest path connecting pixels (white, same as in Figure 5.3c), a four-step refinement yields the final line tracking result (black). (b) A polynomial fit. (c) A Gaussian fit to re-center. (d) A minimum spanning tree through the resulting points in c. (e) Wavelet smoothening.
5.2.3 Temporal consistency checks

The idea is to err on the side of caution: we aim to exclude questionable data from the dataset that we construct for subsequent analysis. As the automated nature of this data analysis makes it straightforward to accumulate vast quantities of data, there is no disadvantage to excluding from analysis the unreliable parts of it.

As a premise, we take the view that when image acquisition is rapid relative to those conformational fluctuations of DNA that occur on distances resolvable within optical resolution, a true line is likely to be similar to those close to it in time, but an unreasonable line is unlikely to satisfy this criterion. Dissimilar lines may arise for a number of uninteresting reasons: mistaken grouping of data in the first level of analysis to identify the starting DNA molecule; molecules whose extension is so limited that it hardly exceeds the optical resolution; molecules out of focus; and an error in generating the minimum spanning tree. Reasoning from this premise, we compare each line, acquired at a given moment in time, to its antecedents and progeny over the span of a few seconds. When two lines are similar, both are regarded valid and included in the final set of lines.

To implement this idea, the line tracing the contour of each DNA molecule is divided into 21 notional fiducial markers of equal spacing along the line and the average distance by which a line is displaced from another line within the temporal vicinity of a few seconds is measured against a threshold value, as drawn schematically in Figure 5.5a. A small average distance means two lines are similar and are likely to both be valid. After this selection, anomalous lines are removed, as illustrated in Figure 5.5b and Figure 5.5c. We find selected lines and subsequent analysis results converge in a wide threshold
window between 0.4-0.8 µm. We typically choose a threshold value of 0.4 µm to insist on high accuracy. A more stringent threshold sometimes fails to capture the dynamics in conformation change and might introduce bias in the dataset.

We find that precision of the final set of lines is improved by performing a series of pre-screenings before the selection. This automated process searches for many anomalous features; when any of them is identified, that frame of the dataset is excluded from analysis. First, we remove regions where line length fluctuation with time is frequent and unreasonably drastic. We compare the fluctuation before and after applying a third-order Savitzky-Golay FIR smoothing filter. Frames associated with large discrepancy (1.3 µm) are excluded. Second, lines that are too short relative to their neighbors in time are considered to be the likely result of partial features due to incorrect grouping of molecules or else to low intensity parts not identified as signal; as these anomalously short lines can potentially introduce large errors in further quantitative analysis, they are removed. Typically we remove data where the length is at least 1 µm less than those from adjacent frames. We find that the most rapid length fluctuations of our physical system occur about one order of magnitude slower than the criterion we use here to reject potentially wrong data. Third, we exclude dim molecules that have both low shape anisotropy and large short axis component as they are likely to be out of focus. Fourth, the size of a candidate molecule is used to further exclude partial features. Fifth, the end-to-end distance of a line is considered, relative to the length of that line; if the ends appear to loop together too closely, as can happen when the image area is close to the optical diffraction limit, this also is considered likely to be anomalous, and is excluded from further analysis.

It is true that in principle, the exclusion of data might risk biasing the
Figure 5.5: Tests of temporal consistency: selecting reasonable lines from temporal comparison of images at different times. (a) Each line is overlaid with 21 notional fiducial points. Within the temporal vicinity of a few seconds, the average distance between each point on any two lines is evaluated (red lines). When this is less than a threshold designation of similarity, the lines are deemed valid and kept. Lines that fail this similarity criterion are rejected as physically unreasonable. In this typical example, the threshold is 0.4 µm. (b) Raw data: three lines from the level of analysis in Figure 5.4. (c) Outcome of the test for temporal consistency: line 2 is rejected. The text describes additional pre-screenings included in these checks.
dataset. Checking the radius of gyration before and after this selection, we find no bias towards a subpopulation of molecules. The distribution of radius of gyration remains unchanged indicating that the criteria, by which lines are considered to be unreasonable, do not depend on size of the molecule. The exception is that the smallest 10% of molecules (radius of gyration smaller than the diffraction limit, which corresponds to less than half the mean radius of gyration of \( \lambda \)-DNA) do not contribute to the final set of lines. Line tracking should not be expected to work well in this situation. In addition, features so small are likely to be either noise or fragmented molecules, and in this respect it is proper to exclude them. Testing directly for hypothetical bias of the data, we also measured average DNA velocity along the electric field direction, with and without the selection of data just described, and found no meaningful difference.

5.3 Application to specific systems

To illustrate how automated line tracking allows visualization of polymer conformation changes in dynamic processes, we now present two examples, both of which will be fully explored in subsequent reports from this laboratory. The point of these examples is simply to illustrate the quality of data that can be obtained routinely, with large statistics and high fidelity, using the image analysis methods presented in this paper.

Figure 5.6 illustrates experiments in which a \( \lambda \)-DNA molecule, attached at one end to the agarose gel in which it is embedded, is subjected to repetitive stretch and release by applying a periodic electric field. In Figure 5.6a, which overlays lines tracked from hundreds of frames when this molecule was subjected to repeated stretching and releasing from a square wave, one
notices that stretch-retraction events differ subtly from one another, differing not just in times for these processes to be accomplished, but also in the paths by which the molecule is threaded through the gel. Figure 5.6b illustrates a small number of the accompanying length fluctuations during these stretch and release events. Whereas stretching and recoiling both transpire on the

Figure 5.6: An example of applying automated line tracking analysis. A λ-DNA molecule with one end attached to the agarose gel in which it is embedded is subjected to repeated stretch and release using a square wave electric field alternating with period about 10 sec between 16 and 0 V/cm. (a) Overlay of all lines tracked over the time of 25 seconds, at 33 frames per second, showing the distribution of paths by which the molecule threads through the gel. Color bar denotes time. Scale bar: 1 µm. (b) Lower panel: Square-wave electric field applied to the DNA over 25 sec. Upper panel: Length fluctuation of lines tracked during this time window, showing stretch and retraction. Vertical dashed lines show the time at which the electric field switches on and off.
timescale of one second, a recoil process is also clearly evident between two stretches, before the molecule takes on a different path.

Figure 5.7 illustrates data from a different experiment: λ-DNA migrating through agarose gel under a DC field of 16 V/cm. This set of data, consisting of \( \sim 4 \times 10^4 \) evaluations of the curvilinear end-to-end distance of the molecule, shows clearly the statistical nature of the curvilinear end-to-end distance: while the mean value is well-defined, and the most probable value is well defined, the distribution around averages is large.

Figure 5.7: A second example of applying automated line tracking analysis. This figure shows a histogram of curvilinear end-to-end distance defined by line tracking, as λ-DNA migrates through agarose gel under a DC electric field of 16 V/cm. The dataset was 30 movies, each of them consisting of 4,000 frames per movie acquired at 33 fps. Number of observations is plotted against length with bin size 0.2 \( \mu \text{m} \).
5.4 Discussion

In this computer age, with large computing power and digital storage capacity readily accessible, one can use inexpensive personal computers to facilitate image analysis of optical images. Here we have introduced automated line tracking methods applicable to tracing the linear coarse-grained shapes of biomolecules, those shapes larger than the optical diffraction limit, and have illustrated their application to analyzing the conformations of fluorescent-labeled λ-DNA when it is stretched by electric fields. These automated methods enable the facile acquisition of large datasets and by rational extension should readily be adaptable to analysis of data obtained from other visualization methods. It is different in spirit from principal component analysis of DNA, which expresses dynamic information in a virtual phase space of orthogonal basis sets that can be problematical to interpret physically [3].

While the fidelity of tracking reported here is believed to have been optimized within the limits achievable using optical resolution, it is certainly the case that this image analysis is limited in resolution. The line tracking introduced in this study represents a coarse-grained representation of the actual contour of λ-DNA; this is why, for example, even the longest lengths plotted in Figure 5.7 are a factor of 3 small than the known contour length of the molecule, 16 μm. Thus, while these methods are well adapted for quantifying time scales of dynamic processes (illustrated in Figure 5.6) and also their distributions (illustrated in Figure 5.7), the numerical values of the lines do not, at the present time, have one-to-one correspondence with the actual molecular makeup, but should be viewed instead as coarse-grained representations. We remark that this coarse-graining can carry physical meanings, especially when the optical resolution limit coincides with the fundamental
length scale of the system; for example, the pore size of agarose gel, \(\sim 200\) nm, nearly coincides with the diffraction limit in this study. Provocatively, the same also holds for many other biomacromolecular solutions.

5.5 Conclusions

We describe a straightforward, automated line tracking method to visualize within optical resolution the contour of linear macromolecules as they rearrange shape as a function of time by Brownian diffusion and under external fields such as electrophoresis. Three sequential stages of analysis underpin this method: first, “feature finding” to discriminate signal from noise; second, “line tracking” to approximate those shapes as lines; third, “temporal consistency check” to discriminate reasonable from unreasonable fitted conformations in the time domain. The automated nature of this data analysis makes it straightforward to accumulate vast quantities of data while excluding the unreliable parts of it. We implement the analysis on fluorescence images of \(\lambda\)-DNA molecules in agarose gel to demonstrate its capability to produce large datasets for subsequent statistical analysis.

5.6 Materials and Methods

The image analysis presented below was performed on data acquired in the following manner.

5.6.1 Fluorescence microscopy

Data were acquired in epifluorescence mode, typically at a frame rate of 33 fps. A 532 nm excitation laser was focused at the rear focal point of an
oil immersion objective (Zeiss, α-Plan Fluor 100×, NA = 1.45) with 2.5× post-magnification to image with a resolution of 64 nm × 64 nm per pixel. Fluorescence images were collected through the same objective and detected by a back-illuminated electron multiplying charge-coupled device (EMCCD) camera (Andor iXon DV-897 BV) after filtering out light from the excitation laser. The movies were converted into digital format and analyzed. A typical dataset consists of 30 movies, each of them consisting of 4,000 frames per movie acquired at 33 fps. The resulting dataset of conformations typically amounts to >104 from thousands of molecules.

5.6.2 DNA samples

λ-DNA (48.5 kbp, Promega) was labeled by covalently attaching dye, a RhB derivative (Mirus) to heteroatoms on DNA, at a labeling density of roughly one dye per 5 base pairs. Single-molecule measurements of DNA chain conformations were made in a miniature gel electrophoresis setup using agarose gel (final concentration 1.5% (w/v)), in 0.5× TBE buffer (45 mM Tris, 45 mM borate, 1 mM EDTA), the DNA being at picomolar concentration. Anti-photobleaching agent, ascorbic acid (SigmaAldrich), was present at a final concentration of 10 mM. A DC voltage was applied across two Pt electrodes to generate an electric field ranging from 6 to 16 V/cm.
CHAPTER 6
DIRECT SEGMENT TRACKING
THROUGH MODULAR STITCHING TO IMAGE SUB-MOLECULAR DYNAMICS

6.1 Background

Almost a generation since Chu et al. first imaged fluorescently labeled single DNA molecules [70], it remains challenging to resolve the submolecular dynamics of a flexible polymeric chain especially when the chain is highly coiled or looped. This limits our understanding of macromolecular transport, as motion at the level of chain segments is a fundamental measure of polymer dynamics [5, 6, 27, 78]. To make progress, one needs to overcome three resolution challenges: the optical diffraction limit, tracking rapid motion, and knowing the relative positions of labeled segment along the chain. While the first can involve localization methods that have driven recent progress in super-resolution fluorescence imaging [79–81], here we approach the latter two with modular assembly of labeled short segments into longer chains such that desired segments of a chain can be labeled with their positions resolved below the diffraction limit and monitored rapidly with respect to the main chain contour.

This chapter is partially reproduced from Guan, J; Wang, B; Bae, S. C.; Granick, S. Modular stitching to image single-molecule DNA transport. J. Am. Chem. Soc. 2013, 135, 6006-6009.
6.2 Protocols and design rules of modular stitching

Our synthesis starts by constructing through restriction enzyme digestion molecular modules with defined overhanging ends. Typically a few kilobases long, they are then covalently labeled with dye and grafted to desired positions of a long parent chain with synthetic “stitching oligomers” through complementary base pairing. As hundreds of overhang sequences can be created by restriction enzymes and oligomers of any desired sequence can be synthesized, the method can potentially label segments at any desired position along a chain and even in more complex molecular architecture and networks through programmable and orthogonal modular stitching. Examples of chain architectures that we anticipate are sketched in Figure 6.1a, with one example implemented below.

The synthesis scheme meets imaging-specific desiderata, although molecular cloning and DNA nanotechnology build on similar ideas but use fewer fluorescent dyes: a) The parent chain size should exceed the diffraction limit (~300 nm or ~10 kbp), just as for many earlier studies in prior literature [2, 70, 82–86]; b) Size of labeled modular segments should be less than the diffraction limit, enabling their localization with nm precision, yet long enough to accommodate sufficient dyes to provide adequate signal-to-noise for imaging in bulk microenvironments over long time (DNA chains of 2500-5000 bps are found empirically to work well); c) Dye pairs, on the labeled segments and on the main chain contour, should possess minimal spectral overlap for simultaneous 2-color imaging.

The specific example studied in this paper concerns single-molecule imaging of dynamics internal to polymer chains: the ends and main chain contour of linear DNA. We constructed a hybrid chain assembled from λ-DNA as the
parent chain, \( \sim 16 \, \mu m \) (or 48 kbp) in length, and two segments (\( \sim 5 \, \text{kbp} \)), each grafted onto an end of the parent chain. Chain ends are commonly believed to play critical roles in polymer dynamics in both equilibrium and driven systems [5, 6, 27, 78], and our hybrid chain allows tracking simultane-

Figure 6.1: Scheme of submolecular modular labeling of DNA. (a) Micron-scale DNA constructs may be synthesized with desired chain architectures and modular labeling. These sketches show a linear chain labeled at designated spots, the ends and within the chain; a 3-arm end-labeled star; a network labeled at the crosslinks. (b-d) End-labeling procedure employed in this work: (b) First, onto the overhanging end of \( \lambda \)-DNA, labeled A (48 kbp) is attached a shorter segment, labeled B (40 bps). The new end overhangs now bear the B sequence, which is available for further reaction; (c) Next, molecules bearing overhang sequence A are purified and labeled with covalently-attached fluorescent dyes; (d) Finally, these DNA fragments in (c) are attached onto the two ends of the DNA prepared in (b) and the chain backbone is labeled with dye molecules of different color.
ously dynamics of chain ends with respect to the main chain. The detailed procedure is now summarized. First, one end of λ-DNA (denoted as A in Figure 6.1b) is modified to bear the same sequence as the other end (denoted as B in Figure 6.1b, an overhang of 12 bases in length) by attaching a short stitching DNA oligomer (~40 bps) to end A and removing excess oligomers through a centrifuge column (Amicon column, Millipore). Meanwhile, we separate λ-DNA HindIII fragments (NEB) into bands using agarose gel electrophoresis and subsequently extract the DNA fragment containing end A from this gel (Qiagen gel extraction spin column) as HindIII generates fragment of desired length and overhang. We prepare molecular modules, kilobases long, from fragments cut by restriction enzymes because synthetic DNA oligomers are available commercially only below ~100 bps. We covalently label each extracted fragment with ~500 Cy5-derivative dye molecules (Mirus Bio) and further concentrate the solution to ~100 nM (Amicon spin column) (Figure 6.1c). Because the ends A and B of λ-DNA are complementary in sequence, we then hybridize the labeled fragment bearing end A with modified λ-DNA bearing end B for 1 day to improve hybridization efficiency (Tris 10 mM, EDTA 1 mM, MgCl$_2$ 10 mM, PH=7.5). Last, we label the hybrid chain with diffusive DNA dyes (Sybr Gold, Invitrogen) in order to visualize the main chain (Figure 6.1d). For this study we ligate the modules to enhance stability of end-attachment before dispersing hybrid chains into molten agarose gel at high temperature.

To boost reaction yield, we found the following strategies useful: a) small stitching DNA oligomers are added in 100-fold excess and later removed prior to hybridization between modules; b) labeled modules are concentrated to ~100 nM to improve hybridization efficiency. Overall we achieve ~30% yield. It is worth noting alternative strategies that failed. For example,
shorter DNA segments (∼100 bps) attached to chain ends could not be imaged faithfully. Although this lower density of dye labeling is appropriate to DNA bar-coding and imaging static conformations on surfaces and in nanochannels [87, 88], imaging rapid motion far from a surface demands brighter dye-labeling.

6.3 Results

To demonstrate the power of this approach, we imaged such chains in gel electrophoresis with emphasis on internal DNA dynamics during this process. The transport features we observe were neither detectable with a conventional whole-chain labeling approach, nor predicted by prevailing theories and simulations. The λ-DNA, labeled as described above, was embedded within 1.5 wt% agarose gel (Fisher, molecular biology grade, low EEO) in the presence of 1× TBE and glucose oxidase-based anti-photobleaching buffer, and visualized using a homebuilt setup for two-color epifluorescence tracking (Zeiss observer.Z1, 100× objective) with time resolution of 50 ms (Andor iXon EM-CCD cameras). Qualitative inspection of the raw data obtained at 9 V/cm (Figure 6.2a) shows that chain extension fluctuates considerably on the time scale of seconds. Rapid data acquisition is needed to see it. Such images were quantified with 10 nm resolution using homebuilt software based on standard methods [56]. Simultaneously tracking both chain ends (red) and the main chain contour, we find that one end of a chain usually (>90% of the time) leads the rest of the chain whereas the other end trails. This preferential end protrusion contrasts with the common expectations of so-called “hernia” or “hairpin” formations which predict random protrusion of any segment to form folded chain conformation [15, 28, 29, 37, 39, 40, 89–92].
Figure 6.2: Illustrations of how end-labeled λ-DNA migrates through 1.5 wt% agarose gel at 9 V/cm. (a) Images of the same molecule during a time span of 10 s. The ends, marked with red crosses, are overlaid onto the chain contour (white). One end of a chain usually leads in migration. Electric force ($F$) points to the right. Scale bar: 1 µm. (b) An illustrative molecule's time-dependent position along electric force direction showing asymmetry between two ends. The leading end (black) advances relatively continuously while the trailing end (red) displays longer pauses and jerky advance. The first few pausing events of leading ends are marked by dashed lines (orange). (c) The extension between two ends fluctuates, reflecting leading end stretch (†) and trailing end recoil (↓). (d) The trajectories of the two ends, leading end (grey) and trailing end (pink) are overlaid for a representative molecule. The respective pause positions are compared, leading end (black) and trailing end (red). The two ends are seen to pause at nearly the same positions within the agarose gel, indicated by arrows. Electric force points to the right. Scale bar: 1 µm. Inset: magnified view of local motion of an end during pause. Region is 200×200 nm².

83
Plotting against time the end position in the electric force direction (Figure 6.2b), we also observe frequent pausing, even though a constant electric force is applied to the chain. Beyond intermittent motion of the center-of-mass motion, we also observe asymmetry between the two ends: while the leading end moves relatively continuously, the trailing end displays long pauses followed by rapid catch-up jumps. Plotting the time-dependent distance between head and tail of a representative molecule (Figure 6.2c), one sees the origin of the large scale chain length fluctuation: stretch of the leading end causes chain extension, and recoil of trailing end causes contraction. Onset of stretch of the leading end ($\uparrow$) and recoil of the trailing end ($\downarrow$) are also highlighted in this figure. A larger dataset is summarized in Figure 6.3.

Inspecting a representative trajectory more closely, one observes that the leading and trailing ends of the molecule follow almost the same migration path through the agarose gel and pause at almost the same locations within the gel, but at different times. This is illustrated in Figure 6.2d, where trajectories of ends are projected into the x-y plane and electric force pointing to the right. Trajectories of the leading end (grey) and trailing end (pink) are seen to pause at nearly the same positions, as denoted by the black points (leading end) and red points (trailing end). Upon surveying multiple molecules, we find that the two ends consistently share ~80% of the pausing locations across a range of electric field. The overlap of pausing locations between chain ends suggests the middle chain segments also pause at these locations although we do not explicitly track the middle segments here. In other words, the resistance to transport is experienced sequentially along a migrating chain as each segment sequentially encounters the same “trap” in the network. Yet there seems to be no fixed adhesion to gel fibers; chain ends remain mobile locally over lengths consistent with estimated mesh size.
Physically, it seems that pausing reflects tight constrictions in the static network structure that hinder transport as a chain squeezes through.

Experiments were also performed with DNA labeled at one sole end. That end-labeling caused neither preferential end protrusion nor intermittency followed from our observation that the labeled ends can either lead or trail.

Figure 6.3: Electric field dependence of asymmetry between two ends quantified after averaging ~100 events involving ~20 molecules at each field. (a) Ensemble-averaged velocity between pauses in the electric force direction is plotted against field $E$ for leading end during stretch events ($\circ$) and trailing end during recoil events ($\triangle$). Error bars are estimated by the standard bootstrap methods. Lines are guides to the eye. (b) For $E = 9\,\text{V/cm}$, the velocity distribution is compared for leading (open bar) and trailing (shaded bar) ends. Velocity is evaluated over 0.5 s. (c) Pause time between transport (stretch or recoil) phases is plotted against $E$. Symbols are same as in panel a. Pause is identified from apparent plateau in time-dependent displacement such as those shown in Figure 6.2b. (d) For $E = 9\,\text{V/cm}$, the pause time distribution is compared. Symbols are the same as in (b).
while displaying similar intermittency and stretch-recoil events. Earlier studies had shown that gel electrophoresis may involve intermittent motion and chain length fluctuation in experiments [72, 93, 94] and early simulations [27, 78, 95]. The novelty of this study is to reveal internal details of the constituent chain dynamics that were, to our knowledge, not identified previously. For example: intermittency was not previously so clearly attributable to passage through tight constrictions in the gel; chain length fluctuation was not identified as stretch of leading end and recoil of trailing end.

Furthermore, submolecular labeling enables one to measure the average moving velocity of leading end during stretch and trailing end during recoil (Figure 6.3a), with stretch and recoil phases readily identified as time between pausing events from time-dependent positions (Figure 6.2b). Figure 6.3 compares the dynamics between two ends after averaging many events involving multiple molecules under electric field spanning from low to high. At low field (6 V/cm) DNA chain conformation is close to that in the absence of field, and never shows large extensions, whereas with increasing field strength, larger fluctuations of chain extension are observed. Although both ends move faster with increasing electric field, consistent with previous studies regarding the center-of-mass mobility of such molecules, recoil of trailing end is faster than stretch of leading end at high electric field (Figure 6.3a). The histogram of instantaneous velocity (Figure 6.3b; evaluated at 9 V/cm) further shows the trailing end has a large portion of fast recoil events. The faster recoil of trailing ends, relative to stretch of leading ends at the highest field strengths, suggests that intra-chain tension facilitates recoil by pulling trailing ends after chain is extended. At low electric field, this difference is less prominent as when the field is low chains are rarely stretched to a high extension and therefore are unable to generate high tension.
Comparing the pause times under different electric field, we see that it decreases rapidly for both ends with increasing electric field (Figure 6.3c). Since protrusion of chain ends initiates chain transport (Figure 6.2a) and the pause duration of leading ends decreases with increasing field (Figure 6.3c), electric field can promote transport by facilitating protrusion of chain ends. This is consistent with the observation of increased fraction of “hooking” events with increasing field (Figure 6.4) which also suggests easier end protrusion at higher field. The pause time of trailing ends exceeds that of leading ends (Figure 6.3c) and is more broadly distributed in the range 0.1-10 sec as the chain gets caught in elongated conformations (Figure 6.3d). The prevailing pattern is that one end of a chain tends to stretch out and pulls slack from the still-quiescent remainder of the chain until the other end is yanked forward.

Figure 6.4: Hooking identified by submolecular imaging. (a) End positions are plotted against time during a hooking event. Comparing the leading (black) and trailing (red) ends of the molecule, one observes a dip for the trailing end denoting backwards motion. (b) The incidence of hooking is plotted against electric field strength.
6.4 Discussion

6.4.1 testing hooking hypothesis

In hooking (or geometration) models, intermittency of transport and chain length fluctuation are considered to reflect chain hooking onto a gel fiber, becoming stretched by this event, and subsequently sliding off it. But we quantified the incidence of hooking by spotting characteristic backwards motion of the trailing end before recoil, akin to motion past a pulley as sketched schematically in (b), and it emerges that hooking is seen in 10% at 6 V/cm, 30% at 9 V/cm, and <50% at 12 V/cm of the total number of recoil events (Figure 6.4). Most recoil events, even when hooking is observed, seem to involve the pausing mechanism discussed in Figure 6.2d. Parenthetically, increased fraction of “hooking” events with increasing electric field is consistent with decreasing pausing duration (Figure 6.3c), both of which suggest easier end protrusion at higher and higher field strength, as increased hooking fraction implies enhanced chance for both ends to protrude simultaneously from different sides of a gel hindrance.

6.4.2 Connection to proposed microscopic pictures

To put this work into context: while macroscopic aspects of electrophoresis (separation efficiency and its dependence on electric field and gel type) have been understood for a long time, the present new observations of microscopic chain dynamics at the single-molecule level encourage us to compare to the prevailing models on internal chain dynamics. Given the chain length and gel mesh size (∼200 nm for this agarose concentration), this system is in the long-chain regime where at equilibrium the mass of a chain spans approxi-
mately 20 meshes. The classical theories of “reptation” biased by an electric field [27, 78] predict neither intermittent dis-placement nor the large chain length fluctuations that we observe. In “hooking” (or “geometration”) models, intermittency of transport and chain length fluctuation are considered to reflect chain hooking onto a gel fiber, becoming stretched by this event, and subsequently sliding off it. But we quantified the incidence of hooking (Figure 6.4), and it emerges that hooking is only seen for a fraction of recoil events (10% at 6 V/cm, 30% at 9 V/cm, and <50% at 12 V/cm). Moreover, even when hooking is observed, chain transport seems to also involve the pausing mechanism discussed in Figure 6.2d. The entropic trapping model does admit intermittency, but it appears to predict random protrusion of all segments [15, 28, 29, 37, 39, 40, 89–92] whereas we observe preferential protrusion at the chain ends. Predictions of the entropic barrier model are best developed for cases of single barrier crossing [15, 28, 29, 37, 39, 40, 89–92], but this data suggests instead a scenario involving multiple barrier-crossing events as segments along the chain sequentially encounter a given tight constriction in the gel network (Figure 6.2d). It would be interesting to revisit the simulation models [27, 29, 78, 96–98] in light of the new findings reported here.

6.4.3 Extension of the modular assembly method

These novel observations are more pertinent to polymer dynamics, than the separation technology of gel electrophoresis, which as an analytical technique already works well [27, 78]. It is a grand challenge in polymer science to understand the motions and relaxations of individual polymer chains [5, 6], and DNA has long been appreciated as a model “polymer” for this purpose.
The methods of submolecular fluorescence imaging, using the facile modular stitching approach described here, can provide new data to which prevailing models can be compared critically. Although the present data is limited so far to just a single DNA chain length (λ-DNA) and to a single agarose concentration, the characteristic transport features would not have been detectable by the conventional whole-chain labeling approach.

6.5 Conclusions

For study of time-dependent conformation, all previous single-molecule imaging studies of polymer transport involve fluorescence labeling uniformly along the chain, which suffers from limited resolution due to the diffraction limit. Here we demonstrate the concept of submolecular single-molecule imaging with DNA chains assembled from DNA fragments such that a chain is labeled at designated spots with covalently-attached fluorescent dyes and the chain backbone with dyes of different color. High density of dyes ensures good signal-to-noise ratio to localize the designated spots in real time with nanometer precision and prevents significant photobleaching for long-time tracking purposes. To demonstrate usefulness of this approach, we image electrophoretic transport of λ-DNA through agarose gels. The unexpected pattern is observed that one end of each molecule tends to stretch out in the electric field while the other end remains quiescent for some time before it snaps forward and the stretch-recoil cycle repeats. These features are neither predicted by prevailing theories of electrophoresis mechanism nor detectable by conventional whole-chain labeling methods, which demonstrate pragmatically the usefulness of modular stitching to reveal internal chain dynamics of single molecules.
6.6 Materials and Methods

6.6.1 DNA sequences

λ-DNA has two overhanging ends that are complementary: 5’GGGCGGCGACCT3’ and 5’AGGTCGCCGCCC3’. The “stitching oligomer” has two strands that are annealed together. One strand is 5’GGGCGGCGACTATAATACCTACCATGATCTAGTTCAAGTACAG 3’ and the other strand is 5’GGGCGGCGACCTCTGTACTTGAACTAGATCATGGTAGTATTAT3’. The overhang sequence of the labeled module for end-labeling is 5’AGGTCGCCGCCC3’.

6.6.2 Miniature gel electrophoresis cell for imaging

A miniature gel electrophoresis setup was constructed to allow direct imaging in epi-fluorescence mode. Test chains were added to agarose suspension at 45 °C at picomolar concentration for single molecule imaging. 1 ml solution was cast into an imaging sample chamber (Lab-Tek, bottom area 20 mm×20 mm) at 45 °C and allowed to solidify at room temperature for 20 min. Both sides of the gel pad were cut and removed to accommodate two Au electrodes running parallel along the gel sides. The Au electrodes were made by electron beam deposition of a 200 nm Au layer onto the glass substrate. The solidified agarose gel slab was immersed in 3 ml buffer in a so-called “submarine” mode. Electrophoresis buffer (1× TBE) was present to ensure electrical circulation and prevent electrolysis. A mixture of glucose, glucose oxidase, calatase and trolox was used to minimize photobleaching. The sample chamber was mounted onto a microscope stage. A DC voltage was applied across the two electrodes through a function generator (Wavetek) to generate
an electric field up to 16 V/cm.
CHAPTER 7

FICKIAN YET NON-GAUSSIAN DYNAMICS IN A SIMPLE HARD-SPHERE COLLOID SUSPENSION

7.1 Background

Evidence is mounting of Fickian yet non-Gaussian Brownian diffusion in multiple systems [55, 56, 99–105]. In other words, even though it is commonly presumed that the random displacements that objects undergo during Brownian motion follow a normal (Gaussian) distribution, it has been observed that displacement distribution can be non-Gaussian when the mean square displacement remains linear in time. This is counter-intuitive as it seems to contradict the predicted Gaussian behavior at long times but has been reported repeatedly in independent systems: particles diffusing on phospholipid tubules [56], particles diffusing in entangled actin [56], liposomes diffusing in entangled actin [55], polymer chains diffusing on surface [99, 100], particles diffusing among swimming cells [101], etc. Some theoretical rationale has been provided that heterogeneity not sufficiently averaged out on short time and length scale might underpin this behavior [56]. But a limitation of the experimental studies is that they concerned unusual systems with complicated specific interactions, making their generality difficult to assess.

This chapter is partially reproduced from Guan, J; Wang, B; Granick, S. Even hard-sphere colloidal suspensions display Fickian yet non-Gaussian diffusion. ACS Nano 2014, 8, 3331-3336.
Seeking to test the idea in a system more generally representative, we have designed the following experiments involving colloidal suspensions, with several considerations in mind. First, as specific interactions can introduce heterogeneity and interfere with diffusion, we sought to achieve hard-sphere interactions. Perfect hard sphere behavior is now recognized impossible to fully realize in the laboratory but the system we selected appears to come as close to this as can be done [106, 107]. Second, to emulate the locally-varying microenvironments that have been hypothesized to underpin this behavior [55], we sought to work in suspensions whose concentration could be varied over a wide range. To implement this, we track with nm resolution the trajectories of colloidal-sized tracer particles embedded in a suspension of larger matrix particles (Figure 7.1a). The particle-particle interactions are simply hard sphere and hydrodynamic. The size ratio of 1:8 allows smaller particles to diffuse through the interstices between larger particles, no matter how closely the larger ones are packed. Evaluating displacement distribution over three decades in probability, we show non-Gaussian behavior while the diffusion is Fickian. Hydrodynamic interactions have been proposed to differ from spot to spot in crowded colloidal environments [108, 109] which might present heterogeneity to moving particles as they diffuse and lead to non-Gaussian displacement, although the underlying mechanism would differ from conventional glassy systems with dynamic heterogeneity.

Diffusion of this sort is common in nature and technology. Examples include water diffusing through sand beds (a geology problem) [111], solvents diffusing through polymer gels (a materials problem) [112], carbon diffusing through steel (a metallurgy problem) [113]. A vast theoretical and experimental literature on diffusion and hydrodynamics focuses on the effective diffusivity [114–123]. To the best of our knowledge, these prior studies did
Figure 7.1: The experimental system. (a) Schematic illustration of small fluorescent probe particles diffusing among larger matrix particles, both of them index and density matched to the solvent. The size ratio is 1:8. (b) Mean-squared displacement of the probe plotted against time on log-log scales at matrix volume fraction $\phi = 0, 0.15, 0.30, 0.45, \text{ and } 0.55$, from top to bottom. From the dashed lines, $D$ is inferred from the equation $<\Delta x^2> = 2Dt$. (Inset) The $D$, units of $\mu m^2/s$, are plotted against $\phi$. The red curve, $D = D_0(1 - \phi)^{1.5}$, is a theoretical prediction taking into account size difference between probe and matrix particles [110]. The error bars show the uncertainty in measuring absolute volume fraction due to the possible inter-particle interaction being less than ideal hard sphere. The error bars of the ordinate are less than symbol size. (c) Non-Gaussian parameter, $\alpha = \frac{<\Delta x^4>}{3<\Delta x^2>^2} - 1$, plotted against time for 5 values of $\phi$. (d) Non-Gaussian parameter plotted against particle mobility. The subpopulations of particles are binned based on time-averaged displacement over $= 0.1$ s. Dashed line is guide to eye. “+” is from simulated Gaussian trajectory. 0%, black diamond; 15%, brown square; 30%, green upside-down triangle; 45%, red triangle; 55%, blue circle.
not address the displacement distributions that are of primary concern here.

7.2 Results

Figure 7.1a shows a schematic representation of the experiment. We embedded trace quantities of fluorescent PMMA particles (~0.001% volume fraction, \( \sim 0.28 \) µm diameter labeled with rhodamine dye, purchased from Edinburgh Research & Innovation Ltd.) within a matrix of PMMA particles that were invisible optically (~2.2 µm diameter, same source), suspended in the standard solvent mixture of cyclohexylbromide and decalin to achieve index-matching and density-matching [124]. Seeking to come as close as possible to a fluid with hard-sphere interactions, we included 1 mg·ml\(^{-1}\) tetrapentylammonium chloride to screening residue charges on the particles which gives a Debye length of ~0.1 µm [106, 107]. Fluorescence images were taken in epi-fluorescence mode at room temperature on a Zeiss observer.Z1 microscope with 63× air objective, using an EMCCD camera (Andor iXon) and data acquisition at 20 frames per second. Using home-developed software [74] the resolution of particle position was <30 nm. Five matrix volume fractions (\( \phi \)) were studied: \( \phi = 0, 0.15, 0.30, 0.45, \) and 0.55.

A typical raw dataset amounted to about a million tracer positions as a function of time from \(~15,000\) particles at each \( \phi \). The mean squared displacement (MSD) is shown in Figure 7.1b, plotted on log-log scales against delay time. One sees that initially it is proportional to delay time but the slope lessens slightly starting at \(~1\) s, on the order of a few collisions between tracer and the matrix as the tracer moves through interstitial space between matrix particles, whereas Fickian diffusion where MSD is strictly linear with delay time is observed when the tracer particles diffuse in pure solvent. The
implied transition length is consistent with the surface distance between matrix particles. The diffusion coefficient was calculated as $<\Delta x> = 2Dt$ in the linear regime, where $\Delta x$ is displacement in a given time interval, and the brackets denote time-average and ensemble-average. This was slightly less than in pure solvent (inset of Figure 7.1b), in qualitative agreement with theoretical predictions taking into account the size difference between probe and matrix particles [110].

As this large dataset included information about fluctuations about the averages, next the non-Gaussian parameter was calculated, $\alpha = \frac{<\Delta x^4>}{3<\Delta x^2>^2} - 1$. Plotting this against time (Figure 7.1c), one observes that $\alpha = 0$ within experimental uncertainty in pure solvent; $\alpha > 0$ otherwise; $\alpha$ increases with $\phi$; and $\alpha$ is nearly constant over the time window. It is true that non-Gaussian behavior is seen in glassy and supercooled liquid systems but those systems are fundamentally different because they display a splitting of mobility between different subpopulations, what is called dynamic heterogeneity [59, 60, 126, 127]. Here we observe no such split of mobility (Figure 7.2). We calculated the non-Gaussian parameter for subpopulations separated according to their different mobility; $\alpha$ was the same regardless of mobility of that subpopulation, suggesting that this splitting-mobility hypothesis could not explain the data (Figure 7.1d). Trivially, the $\alpha$ of subpopulations were smaller than those of the ensemble because the former presented narrower displacement distribution. This we confirmed by simulating trajectories with strictly Gaussian statistics. Also, the non-Gaussian parameter in our system has a small value of ~0.1-0.3 whereas for glassy systems it is typically ~1-6, an order of magnitude larger.

This dataset was large enough to evaluate the full displacement distribution, including rare events, evaluated over 3 decades of probability. First,
Checking for dynamic heterogeneity by inspecting the time-averaged mean square displacement (MSD) of individual trajectories, we calculated the ratio of MSD of individual trajectories to the ensemble average, specifically

\[ \xi = \frac{\sigma^2(\Delta, \tau)}{\sigma^2(\Delta, \tau)^{\text{ensemble average}}} \]

where \( \sigma^2(\Delta, \tau) = \frac{1}{\tau - \Delta} \int_0^{\tau - \Delta} [r(t + \Delta) - r(t)]^2 \, dt \), and \( \langle \cdots \rangle \) denotes ensemble average. Here \( \tau = 3.3 \text{ s} \) and \( \Delta = 0.1 \text{ s} \), where \( \tau \) denotes how long each trajectory lasts and is selected to be sufficiently long to ensure the accuracy of \( \xi \) but sufficiently short that the distribution of \( \xi \) can be calculated from many trajectories. The symbol \( \Delta \) is the time interval used to calculate MSD to reflect the dynamic spread. Note the slight tendency for the distribution of \( \xi \) to broaden with increasing volume fraction. Colors indicate \( \phi = 0 \) (black), \( \phi = 0.15 \) (brown), \( \phi = 0.30 \) (green), \( \phi = 0.45 \) (red), \( \phi = 0.55 \) (blue).

Figure 7.2: The distribution of ergodic breaking parameter \( \xi \) [125].

consider obstacles presented by the matrix at \( \phi = 0.45 \). Figure 7.3a shows relative probability plotted logarithmically against displacement, evaluated for various time lags. The full distributions show consistent deviations from Gaussian but lack any tendency to show an exponential tail of displace-
ment distribution, differing from the pattern typical for supercooled liquids, which is Gaussian at small displacements and exponential at large displacements [127, 128]. Strikingly, they collapse with normalized displacement \( \Delta x' = \frac{\Delta x}{\sqrt{2Dt}} \), consistent with their Fickian displacement.

These distributions, broader than Gaussian, are compared for different volume fraction in Figure 7.4a, which compares distributions as a function of the respective \( \Delta x' \) with the area under each curve normalized to unity. In pure solvent, perfect Gaussian behavior was observed. For stricter comparison to Gaussian behavior, the data were normalized to the Gaussian curve. Plotting this ratio against \( \Delta x' \) on a linear scale (Figure 7.4b), one sees more explicitly that discrepancies were most pronounced for the largest displacements, those displacements whose probability was lowest. It is noteworthy that the displacement distributions deviate from Gaussian behavior already at \( \phi = \)

![Figure 7.3: How the displacement probability distribution scales with time. (a) Displacement probability distributions over 3 orders of magnitude plotted logarithmically against displacement, \( \Delta x \), with ordinate normalized to the maximum, at volume fraction 0.45, at times 0.1, 0.2, and 0.3 s. Dashed line is Gaussian fit to the small displacements at 0.1 s. (b) The same ordinate as left, plotted against \( \frac{\Delta x}{\sqrt{2Dt}} \), where \( D \) is diffusivity from slope of mean-squared displacement versus time and area under each curve is normalized to unity. Note the collapse of the data.](image-url)
0.15, which traditionally would be considered dilute. Given that the matrix may crystallize at \( \phi = 0.55 \) [129, 130], one might expect this to interfere, but

Figure 7.4: Dependence on volume fraction. (upper) Displacement distributions plotted logarithmically against displacement normalized by \( \sqrt{2D\Delta t} \) at \( \Delta t = 0.2 \) s for different volume fractions of matrix particles: \( \phi = 0 \) (black), 0.15 (brown), 0.30 (green), 0.45 (red), and 0.55 (blue), compared to Gaussian distribution (orange line), \( \sqrt{2\pi} \exp^{-x^2/2} \). Area under each curve is normalized to 1. (lower) For each volume fraction, the ratio of the observed probability distribution to the Gaussian is plotted against normalized displacement.
we see no significant change in nanoparticle diffusivity nor a split of mobility into fast and slow subpopulations at this concentration (Figure 7.2). One possible explanation is that the actual concentration is less than this; it is reported that the absolute value of \( \phi \) can shift \( \pm 3-6\% \) in this system [106, 107]. Regardless of why, this volume fraction presents simply a systematic extension of tendencies already apparent when the matrix concentration is less. Further, there is no suggestion of lack of equilibration, as neither aging nor change of nanoparticle mobility are seen over the course of the observation (Figure 7.5). Deviations from Gaussian are more prominent with increasing \( \phi \), however. Note also the curious feature that all curves appear to intersect at the same two values of \( \Delta x' \) although no interpretation of this empirical observation is offered at this time.

Physically, the greater heterogeneity suggested by the broader distribution of displacement could reflect local obstacle concentrations that differ from the average \( \phi \) according to the local configurations of matrix particles. This is why, reflecting locally relatively crowded and sparse obstacles, we would observe more small displacements and large displacements than Gaussian; the fewer intermediate-sized displacements would then be a consequence of normalizing area of distribution to unity. These trends become more pronounced with increasing \( \phi \). Such heterogeneity was supported by further analysis. First, individual trajectories were inspected. Their time-averaged mean square displacement was also found to become increasingly heterogeneous with \( \phi \) (Figure 7.2). Second, adjacent steps in time were found to be correlated in magnitude; that is, large steps were likely to be followed by large steps and vice-versa. This gave a U-shaped conditional displacement magnitude after a given displacement, centering around zero displacement. This trend grew systematically with increasing \( \phi \), whereas in pure solvent
this bias was not observed (Figure 7.6). In the course of this study, we re-analyzed already-published data from this laboratory [55, 56] and found the same qualitative trend of the conditional displacement, though the analysis was not made at the time of those publications about other systems that are Fickian yet non-Gaussian.

Matrix particles diffused slowly in this experiment but were not immobile as in systems with quenched disorder [131, 132]. Then it is reasonable to expect that the smaller tracer particles experienced spatially-varying en-

![Figure 7.5: No aging or change of mobility was observed for the sample at $\phi=0.55$. This sample was pre-equilibrated for 12 hrs on the microscope stage before the experiment, then measurements were made for an additional 8 hrs. The dashed line shows the ensemble-averaged diffusivity.](image-url)
Figure 7.6: Conditional displacement of successive time intervals. The abscissa is displacement in the first time interval. (Upper panel) The ordinate is root mean square displacement in the second time interval. The U-shaped pattern centered around zero displacement, observed in the presence of matrix particles but not in pure solvent, shows that large displacements are likely followed by large displacements in the presence of matrix particles. This trend grows with increasing volume fraction. Time interval is 0.1 s for each $\phi$. (Lower panel) The ordinate, mean displacement in the second time interval, equals zero regardless of volume fraction. Time interval is 0.2 s ($\phi=0$, black), and 0.1 s otherwise. Colors indicate $\phi=0.15$ (brown), $\phi=0.30$ (green), $\phi=0.45$ (red), $\phi=0.55$ (blue).

environments, the latter fluctuating more slowly. Given that tracer particles diffused only a fraction of the matrix particle size over the experimental time window, the local environment experienced by different tracer particles d-
iffered from point to point according to the random arrangement and the local hydrodynamics of the slowly fluctuating matrix particles. Static obstacles on surfaces have shown to produce anomalous dynamics in 2D [104, 105] and the current system shares these features qualitatively in 3D. It seems that the observed non-Gaussian displacement distributions essentially reflect heterogeneity not averaged out on the time and length scales that we study.

7.3 Discussion

Though non-Gaussian diffusion is sometimes identified in supercooled or glassy systems [59, 60, 126–128], the accepted view of glassy behavior seems to differ significantly. In the trajectories of particles in glassy systems, it is common to analyze the persistence and exchange times (here, the persistence time is the first-passage time for a given particle and exchange time is the waiting time for subsequent passage) and it is found that they decouple typically [133]. The persistence and exchange time distributions in this system, plotted against time lag in Figure 7.7 (top), display no decoupling. In fact, these two distributions appear to be strictly identical, as one sees from their ratio plotted against time in Figure 7.7 (bottom).

We now compare our data to the popular continuous time random walk (CTRW) and fractional Brownian motion (FBM) models to describe anomalous diffusion [125, 134]. Recently, velocity correlation functions have been suggested to distinguish these processes. We observe a velocity anticorrelation at short time calculated from frame-based displacement correlation, contrasting to unbounded CTRW predictions that do not show negative values due to the absence of correlation between different jumps (Figure 7.8). It is true that we find certain dynamics features consistent with FBM mod-
els: first, the slight velocity anticorrelation at short time; distribution of the ergodicity breaking parameter centered around 1 which is an indicator of the individual particle mobility (Figure 7.2, Figure 7.8). However, FBM models require a Hurst exponent less than 0.5 to yield velocity anticorrelation which simultaneously leads to subdiffusion with a MSD exponent of less than 1,

![Figure 7.7: Persistence and exchange time analysis. Top panel: distribution of persistence (solid) and exchange (open) times, compared for $\phi = 0$ (black), 0.15 (brown), 0.30 (green), 0.45 (red), and 0.55 (blue). Threshold distance $d$ is set to 1 $\mu$m. No decoupling was seen for $d = 0.5 - 5 \mu m$. Lines are guides to the eye. Bottom panel: ratio of the distribution of persistence to exchange time at each time lag, plotted with a cutoff of $P(\text{log}t) > 0.05$. Dashed line showing a ratio of 1 is shown for comparison.](image-url)
contrasting the Fickian diffusion we characterized here with MSD linearly growing with time. Physically, the assumptions in CTRW or FBM models are not validated in the current system, however. We do not observe the trapping and discrete jumps in dynamics assumed in CTRW models. We do not observe the memory effects assumed in FBM models.

From data in the literature, the sub-diffusive to diffusive transition for matrix particles of this same size at volume fraction 0.45 occurs at \( \sim 100 \, \text{s} \) [59]. Also from literature, the time for a matrix particle to diffuse its own

Figure 7.8: Velocity autocorrelation function \( C_v(\tau) \) plotted against time lag \( \tau \) on linear scales, where \( C_v^{(\varepsilon)} = \frac{1}{\varepsilon^2} < (x(\tau + \varepsilon) - x(\tau))(x(\varepsilon) - x(0)) > \) and here \( x \) is instantaneous nanoparticle position, \( \tau \) is the delay time, and \( \varepsilon = 0.05 \, \text{s} \). A slight anti-correlation at 0.1 s quickly decays to 0. The anti-correlation increases slightly with increasing volume fraction. Colors indicate \( \phi = 0 \) (black), 0.15 (brown), 0.30 (green), 0.45 (red), 0.55 (blue). (Inset) a magnified view of the anti-correlation region.
size at $\phi=0.45$ is estimated to be $\sim 100 \text{ s}$ [60]. We identify this timescale with the longest relaxation timescale of the matrix. Because the matrix particles diffuse so slowly and are obstructed by their neighbors, over our experimental window of $<10 \text{ s}$ the matrix structure would not rearrange.

Inspection of the velocity autocorrelation function implies that collision and back scattering events occur on the time scale $\sim 0.1 \text{ s}$ (Figure 7.8). Analysis of the mean-square displacement implies that the transition time from diffusion to sub-diffusion in this interstitial diffusion process is $\sim 1 \text{ s}$ (Figure 7.1b). This longer time is reasonable physically, as the transition should require several collision and back scattering events. In fact, the implied length scale is calculated to be roughly the average surface distance between matrix particles at the respective volume fractions (Table 7.1).

The novelty of the present experiments is to demonstrate non-Gaussian yet Fickian diffusion in a very simple system, with physical interactions much simpler than the somewhat esoteric previous systems (actin networks, phospholipid tubules, vesicles) in which this pattern was earlier identified [55, 56]. Here the mechanism is probably that hydrodynamics experienced by diffusing particles differs from spot to spot as recognized qualitatively long ago [108, 109] without exploring how this would modify the probability distribution of displacement and how the slowly varying local hydrodynamics due to the slowly fluctuating matrix perturbs diffusion. The degree of heterogeneity characterized here lies between one extreme of the simplest fluid of all (no

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>mean surface distance ($\mu$m)</th>
<th>$t_{\text{collision}}$ (s)</th>
<th>$t_{\text{transit}}$ (s)</th>
<th>$l_{\text{transit}}$ ($\mu$m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15</td>
<td>2</td>
<td>6</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.30</td>
<td>1.1</td>
<td>2</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>0.45</td>
<td>0.7</td>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.55</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 7.1
Figure 7.9: Conditional displacement of the $N$th step given the displacement of the 1st step, with successive time intervals of 0.1 s and $N$ up to a maximum of 10, at volume fraction $\phi=0.45$. The U-shaped pattern flattens slowly with increasing $N$. The horizontal dashed line shows the expected displacement value for zero correlation, calculated from the mean square displacement at this time. The data are plotted for $N=2(\ast), 3(+), 4(\times)$, and 10(○) corresponding to a lag time of 0.1 s, 0.2 s, 0.3 s, and 0.9 s respectively.

matrix obstacles, $\phi=0$) and the other extreme of harsh obstacle obstructions (supercooled liquids and glasses). Our observations of more frequent large steps than the Gaussian assumption can be expected to influence the outcome of events whose essence lies in their rarity: dynamics determined by first-passage time, rare-event initiated chain reactions, diffusion-limited reaction, triggering and signaling. The phenomenology reported here can be
expected to be general as colloidal suspensions are so common.

The new findings presented here underscore the prominence of non-Gaussian diffusion despite linear mean square displacement. While to observe this in our crowded situation of highest volume fraction may be easy to understand physically, we emphasize that the tendency starts even when the volume fraction is so low that it might reasonably be considered dilute and expected to follow Gaussian dynamics. The findings underscore, in agreement with recent studies of more esoteric soft matter systems, the seeming ubiquity of non-Gaussian yet Fickian diffusion. This physical situation was previously considered by much earlier analysis of effective diffusivity [114–123], but this literature did not address theoretically the displacement distributions that are of primary concern here, and the single-particle data presented here was inaccessible to experimentalists at that time. We interpret these results as finding new life in this problem.

7.4 Conclusions

We scrutinize 3 decades of probability density displacement distribution in a simple colloidal suspension with hard sphere interactions. In this index-matched and density-matched solvent, fluorescent tracer nanoparticles diffuse among matrix particles that are eight times larger, at concentrations from dilute to concentrated, over times up to when the tracer diffuses a few times its size. Displacement distributions of tracers, Gaussian in pure solvent, broaden systematically with increasing obstacle density. The onset of non-Gaussian dynamics is seen in even modestly dilute suspensions which traditionally would be assumed to follow classic Gaussian expectation. The findings underscore, in agreement with recent studies of more esoteric soft
matter systems, the seeming ubiquity of non-Gaussian yet Fickian diffusion.

7.5 Materials and Methods

Trace amounts, 0.001% in volume fraction, of fluorescent PMMA particles 0.28 µm in diameter (purchased from Edinburgh Research & Innovation Ltd), were dispersed evenly into non-fluorescent PMMA particles 2.2 µm in diameter, in a suspension medium of index-matched and density-matched mixture of cyclohexylbromide (98%, Aldrich) and decalin (99%, Aldrich), with 1 mg/ml tetrapentylammonium chloride added to screen residual charge between particles. The sample cell was assembled using coverslips as spacer sandwiched between a glass slide and an imaging coverslip. First, the particles were centrifuged in nearly density-matched solvent to achieve random close packing of ~64%, then they were diluted with the density-matched solvent mixture to the desired lesser volume fraction. Samples of 0, 0.15, 0.30, 0.45, and 0.55 in volume fraction were used in this study. The colloidal dispersion was inserted into the sample cell through an inlet hole drilled through the glass slide and rapidly sealed with molten mixture of galactose and dextrose (1:1). The sample was allowed to equilibrate on the microscope stage for a few hours and epifluorescence microscopy imaging was carried out at a focal plane at least 30 µm away from the surface. Spatial resolution of ~30 nm was achieved using a 63× air objective (NA = 0.75) with 1.6× post-magnification and image analysis software written in-house.
REFERENCES


