CHALMERS
UNIVERSITY OF TECHNOLOGY

Developing data analysis methods for Nanofluidic Scattering Microscopy

Master's thesis in Physics

AARÓN DOMENZAIN DEL CASTILLO CERECER

DEPARTMENT OF MICRO AND NANOTECHNOLOGY

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

Master's thesis 2024

Developing data analysis methods for Nanofluidics Scattering Microscopy

AARÓN DOMENZAIN DEL CASTILLO CERECER

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Physics
Envue Technologies
Chalmers University of Technology
Gothenburg, Sweden 2024

Developing data analysis methods for Nanofluidic Scattering Microscopy
AARÓN DOMENZAIN DEL CASTILLO CERECER

Supervisor: David Albinsson (Envue Technologies)
Co-supervisor: Barbora Špačková (Envue Technologies, Institute of Physics of the Czech Academy of Sciences)
Examiner: Christoph Langhammer (Department of Physics, Chalmers University of Technology)

Master's Thesis 2024
Department of Physics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Generated by Adobe Firefly AI image generator. Prompt: Brownian particles diffusing through a long nano channel, shining due to being irradiated with a laser in impressionist art style, seen from above.

Typeset in LaTeX
Gothenburg, Sweden 2024.

Developing data analysis methods for
Nanofluidics Scattering Microscopy
AARÓN DOMENZAIN DEL CASTILLO CERECER
Department of Physics
Chalmers University of Technology

Abstract

Nanofluidic Scattering Microscopy (NSM) is a label-free technique that allows the simultaneous measurement of molecular weight and size of single molecules and nanoparticles flowing through a nanochannel. Assuming free diffusion, the diffusivity coefficient derived from the trajectory of a particle can be utilized to determine its hydrodynamic radius via the Stokes-Einstein relation. However, non-specific atractive interactions between biological nanoparticles and the inner walls of the nanochannel, known as biofouling, immobilize the particle, invalidating the assumption of free diffusion and making impossible to accurately determine the particle size. This study introduces a computational method to detect biofouling in NSM, based on the statistical analysis of instantaneous kinetic energy in discrete Brownian motion. The method sets a probability threshold to either accept or reject the hypothesis of free diffusion, allowing for the exclusion of biofouled segments from diffusivity calculations, thereby enhancing the accuracy of hydrodynamic radius determination.

Keywords: Nanofluidic scattering microscopy, biofouling, Brownian motion, hydrodynamic radius, computational algorithm

Acknowledgements

I am immensely thankful to my main supervisor David Albinsson, for all the support, patience, guidance and horizontality; for our nurturing scientific discussions, advice on programming and endless conversations on bread making, food from the world, language and politics.

To Barbora Špačková and Christoph Langhammer, thank you for your invaluable support and scientific advice. Your understanding and the opportunity to work in such a high-level scientific environment have been a cornerstone in my growth.

To the Envue Technologies team, for your lovely company, showing me the Swedish working culture and all those nice fikas which alleviated my stress during the hard times.

To Amanda and the Jäverfelt family, my Swedish family, thank you for believing in me and giving me a place in your hearts.

To my brother, PhD. Iván Domenzain, the first scientist of the family. For all your love and for showing me the way of science at an early age.

To my parents: Gracias por siempre creer en nosotros y enseñarnos a trabajar duro por nuestros sueños; sin ustedes nada.

Aarón Domenzain, Gothenburg, October 2024

Aim

The aim of this master thesis is to study the physical principles of Nanofluidic Scattering Microscopy (NSM), focusing on Rayleigh scattering theory and Brownian motion. It is intended to identify current challenges in NSM, and determine possible solutions *post experimentum* with new methodologies. Such methods shall be derived from first principles and implemented in computational algorithms in MATLAB to integrate the routines to current NSM software developed and used by Envue Technologies. Lastly, the algorithms should be applied on simulated and real experimental data to evaluate the performance and improvements, and after determining limitations, the algorithms shall be integrated into current data analysis methods applied by a cutting-edge, real-world biotech company.

Contents

1

Introduction

We live in a world where the miniaturization of technologies permeates almost every aspect of human life. Over 6 decades ago, Richard Feynman delivered his renowned lecture titled "There's plenty of room at the bottom" at Caltech [1, 2], in which he imaginatively envisioned the potential of manipulating matter down to atomic scales. Although there has been extensive debate regarding whether Feynman's assertions contributed to the actual development of what is presently known as nanoscience and nanotechnology[3, 4] his prophetic vision resonates each time we use our electronic devices utilized for professional and recreational purposes, the sports apparel crafted from engineered textiles, in the paints, polymers, and plastic coatings found in both industrial and domestic environments, and prominently in the pharmaceutical realm, where a vast collection of specialized tools and methodologies are employed, ranging from the production of basic analgesics to the development of the Nobel Prize-winning mRNA vaccine, which was indispensable in addressing the recent global COVID-19 pandemic [5].

Technologies based on nanoparticles[1] require precise methods to measure particle size and molecular weight, as their physicochemical properties are strongly dependent on particle radius. For instance, the catalytic activity of Au nanoparticles is determined by their particle radii [6], due to their high surface-to-volume ratio compared to its macroscopic counterpart. Similarly, size also plays an important role in the cytotoxicity of most nanoparticles when in contact with biological tissues [7, 8]. In the field of biomedicine, biomarker detection is essential for disease monitoring, early diagnosis, and the evaluation of therapeutic responses [9]. High molecular weight biomarkers, such as ferritin, responsible for iron storage and transport in blood [10], and adiponectin, which is crucial for glucose regulation, lipid metabolism, and insulin sensitivity [11, 12], are commonly examined. In general, the analysis of biomarkers is fundamental to evaluate the general health condition of patients. Since each biomarker has a specific range of sizes and molecular weights, the use of precise and accurate analytical instruments is critical, for which a collection of spectroscopy and microscopy techniques has been adapted and adopted in molecular biomedicine and other related fields.

Over the past decades, innovations in microscopy techniques have significantly driven progress in the elucidation of governing mechanisms of biological events at the cellular level and beyond. A stellar example is the historic breakthrough of

[1] Particles whose typical radii span from 1 to 1000 nm.

Watson and Crick's realization of the 3D double helix DNA structure, significantly aided by Rosalind Franklin and Maurice Wilkins' X-ray crystallography images [13, 14], a technique considered by some as a historic precursor of modern electron microscopy. A decade later, the first visualization of the ribosome was achieved by George E. Palade using electron microscopy [15]. Both discoveries mark significant milestones in our understanding of molecular biology, windows to look at life at its fundamental level, for which they were both awarded Nobel Prizes [16, 17].

Why single-molecule and label-free experimental techniques matter?

In scientific research on nanoparticles and biomolecules, the degree of our understanding is often dependent on the methods we use to study complex systems. Traditional approaches that focus on measuring bulk or average properties can sometimes lead to misleading conclusions, as they overlook the unique behaviors of individual components within a system. This challenge is particularly evident in molecular biology, where single-molecule experiments have emerged as a powerful tool to reveal the intricate dynamics that are hidden in studies of entire populations. By examining individual molecules, these experiments provide a deeper and more accurate insight into the fundamental processes that drive life at its most basic level. Imagine trying to understand the performance of a symphony orchestra by only measuring the average volume of sound produced during a performance. Although this bulk measurement might give you a general idea of the overall intensity, it would completely miss the intricate dynamics of individual instruments, the subtle nuances of solos, and the unique contributions of each musician. Similarly, in molecular biology, studying only the average properties of a large population of molecules can obscure the critical behaviors and interactions of individual molecules. This is why single-molecule experiments are crucial as they allow us to uncover the unique characteristics and behaviors that are often hidden in bulk measurements, providing a more accurate and detailed understanding of biological processes.

This project was carried out in collaboration with Envue Technologies, a start-up company located in Gothenburg, Sweden, which holds the patent for Nanofluidic Scattering Microscopy technology and is actively engaged in its ongoing development.

2

Theoretical background

Nanofluidic Scattering Microscopy (NSM) is a microscopy method[1] that enables the detection of nanoparticles and the simultaneous determination of their individual size and mass, without the need of any labelling technique. In principle, NSM can perform individual detection and measurement of any polarizable nanoparticle. However, in recent decades, there has been a strong interest in single-particle measurements of biomolecules, prominently in the field of nanobiomedicine [18]. For this reason, NSM has been mainly applied to study biological particles, including proteins, lipids, DNA and lipid nanoparticles. The physical foundations and experimental implementation of NSM were elucidated and published by Špačková et al.[19] in 2022. A short description of the setup, as well as the working principles will be outlined in this chapter.

2.1 NSM platform

NSM depends on three major physical components: 1) A nanofluidic chip, 2) A controllable visible light source, and 3) A dark-field microscope equipped with a high speed camera. This physical setup will be portrayed shortly (see Fig. 2.1). Thereafter, we will center our attention on the physics of light scattering. An additional (non physical) component is the custom software for data analysis *a posteriori* [19, Sup. Info.], which will be explained in Chapter 3:Methods.

2.1.1 Nanofluidic chip

The NSM platform consists of a chip based on a homogeneous and optically transparent matrix (usually SiO_2), in which very narrow channels are fabricated. Since the cross-sectional dimensions of the channels fall in the range of 1-1000 nm, they receive the name *nanochannels*. These nanochannels are filled with a fluid (usually water or water-based buffers) and, by controlling the pressure difference between the ends of the nanochannel, and knowing the cross-sectional area, the flow velocity can also be controlled, hence, the name *nanofluidic*. There are numerous methods to fabricate nanofluidic channels [20]. Chemical etching is commonly used due to its practicality, scalability, and cost effectiveness compared to nanolithography methods [21]. However, for the purposes of NSM, reactive ion etching is favored due to the ability to achieve precise patterns and resolutions of tens of nanometers [22].

[1]Since the range of applicability is tipically 1-100 nm, NSM is, strictly speaking, a *nanoscopy* technique.

Maintaining a constant, well-defined cross-sectional area of the channel is desirable for NSM, as we will learn in the following sections. A small volume of a diluted sample containing the biomolecule(s) of interest is administered into the nanofluidic chip for subsequent analysis.

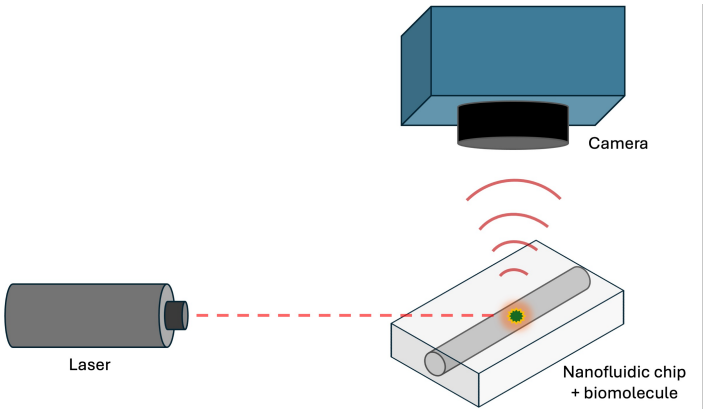


Figure 2.1: Illustration of the NSM setup. A laser beam is pointed to the nanofluidic chip containing the biomolecule to analyze. The system scatters light according to Rayleigh Scattering regime. The scattered light is collected and digitized by camera equipped with an objective placed near the chip. Thereafter, the signal is processed with custom software, yielding to the determination of particle radius and molecular weight of single molecules.

2.1.2 Laser

A collimated light beam is directed at the nanofluidic chip to provoke the nanochannel and biomolecule to scatter light. There is no need for a monochromatic laser, as it will be shown that the working principle of NSM does not depend on the incident wavelength. The light intensity is only controlled to improve the signal-to-noise ratio.

2.1.3 Objective and camera

A dark-field microscope coupled with a high speed[2] CMOS camera is placed in the proximity of the nanofluidic chip. The objective of the microscope amplifies the total scattered light, and the camera captures and records the optical signal as a map of intensity as a function of position and time, known as *kymograph* (see Fig. 3.1a).

2.1.4 Overview of an NSM experiment

Suppose that we have a sample containing biological nanoparticles (BNPs). These samples are usually diluted in a water-based buffer such as PBS [23] with a refractive index similar to water [24]. A small volume of a diluted sample containing

[2]Effective speed of camera is 100-200 frames per second in current NSM experiments. This is fast enough to capture the Brownian motion of biomolecules, yet, each frame capture is long enough to allow a sufficient exposure time.

the BNPs of interest is administered into the nanofluidic chip. The fluid then fills the interior of the channel, so the particles are free to move in the interior of the channel. A laser of visible light is directed towards this optical system. An dark-field microscope equipped with a high speed camera is placed near the chip in such a way that it collects only the light scattered by the system, and not the light coming directly from the laser. Since the wavelength of the visible light is much larger than the characteristic lengths of the particle and the axial dimensions of the nanochannel, the whole system will scatter light according to the Rayleigh scattering regime, which will be explained with more detail in the following section. The total scattered light results from the contributions of both the channel and any particle inside the channel. Since we know the functional dependence of each intensity component constituting the total intensity, we can subtract the effect of the nanochannel from the total overall optical signal, isolating the optical signature coming of the particle only. We will also see that the optical signature of an individual particle contains valuable information which can be used to calculate its size and molecular weight. Since the particles remain in their native state without the need of any modification or selective binding to a surface, NSM is said to be a label-free, single-molecule microscopy technique.

In the following sections, the derivation of a physical model of a nanochannel containing a nanoparticle is presented. This will enable a experimental way to extract the molecular weight and the size of a single particle from a kymograph.

2.2 Rayleigh scattering theory: Nanoparticle inside a nanofluidic channel

We start by assuming a biomolecule as an approximately spherical nanoparticle immersed in a nanochannel, which we consider as an infinitely long[3] cylinder of cross-sectional area A. The dimensions of the particle are much smaller than those of the channel. We model the incident laser beam as an electromagnetic wave with wavelength λ that interacts with the system. The wavelength of the incident light is assumed to be much larger than the characteristic lengths of the system, that is, $\sqrt{A}/\lambda << 1$ and $2\pi R/\lambda << 1$. This implies that both the cylinder and the particle will scatter back the incident radiation following the Rayleigh scattering regime [25–27]. We can interpret scattered radiation as arising from radiating electric and magnetic multipoles composing the scatering body [28]. In this regime, the phase variation of the incoming electromagnetic wave across the surface of the scattering body is negligible, so all the multipoles radiate back with the same phase, which in turn facilitate calculations. The light scattered by a particle and a cylinder will be derived below. Then, we will appropriately combine both effects to construct a model for NSM.

[3]It is obvious that in reality the channel has a finite length, but we consider it practically infinite since it is several orders of magnitude larger than the wavelength λ of the incident light.

2.2.1 Biomolecule as a sphere-like dielectric particle

The power P_p (in W) of the light scattered by a particle is proportional to its cross section σ_p and the intensity of the incident light I_0

$$P_p = cI_0\sigma_p \tag{2.1}$$

with the proportionality constant c being the collection efficiency. The average cross section for linearly dielectric, spherical, uniform particles in the Rayleigh regime can be written as [26, p. 140]:

$$\sigma_p = \frac{k^4}{6\pi}|\alpha_p|^2 \tag{2.2}$$

where $k = \frac{2\pi n}{\lambda}$ the wavevector inside the channel, n the refractive index of the medium within the channel, and α_p the polarizability of the biomolecule. We shall interpret the polarizability of a particle as the propensity of its electron cloud to be distorted due to the interaction with an electric field. This property enables the electric field to interact with the molecule and therefore act on it. Now, the polarizability of an isotropic sphere in a uniform electric field of wavelength much longer than its radius R is given by [28]

$$|\alpha_{sphere}| = 4\pi R^3 \frac{\epsilon_p - \epsilon}{\epsilon_p + 2\epsilon} \tag{2.3}$$

in which the relative electric permitivities and the refractive indices for both the medium and the molecule are related by $\epsilon = n^2$ and $\epsilon_p = n_p^2$, respectively. If we model the biomolecule as an isotropic ellipsoid with semi-axes a, b, c, we can write its polarizability as [26, p. 153]:

$$|\alpha_p| = 4\pi abc \frac{n_p^2 - n^2}{3n^2 + 3L(n_p^2 - n^2)} = V_p \frac{n_p^2 - n^2}{n^2 + L(n_p^2 - n^2)} \tag{2.4}$$

with V_p the volume of the ellipsoidal biomolecule and L being the depolarization factor, that takes into account geometrical effects of the ovoid shape. For a perfect sphere, $a = b = c = R$ and $L = \frac{1}{3}$, which is in complete agreement with Eq. (2.3). For ellipsoids, L ranges from 0 to $\frac{1}{3}$.

By combining Eqs. (2.1), (2.2) and (2.4) we observe that the intensity of light scattered by the biomolecule is quadratic in its volume and quartic in the difference of the squared refractive indices.

2.2.2 Nanochannel as a cylinder

In order to understand the light scattered by the nanochannel inside the nanofluidic chip, we first study the scattering of a long cylinder. We consider a cylinder with cross-sectional area A being irradiated with light of wavelength λ, such as $\sqrt{A}/\lambda << 1$. This situation is well described by Rayleigh scattering theory. The problem of a finite cylinder is not exactly soluble [26, p. 163], but becomes tractable if assumed infinitely long. This assumption is supported by the fact that the length

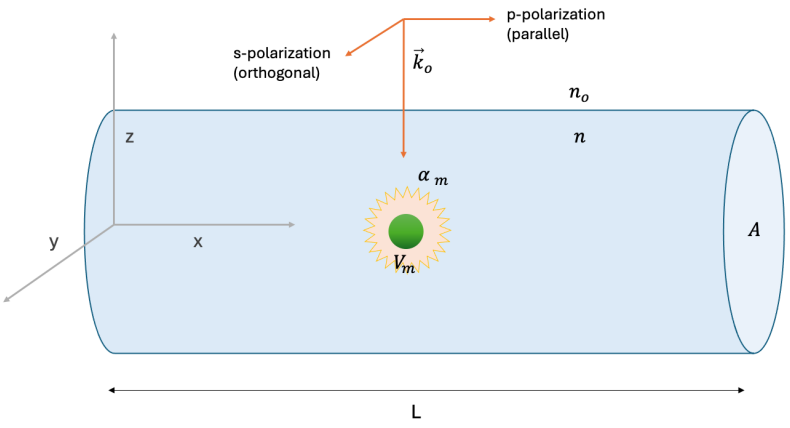


Figure 2.2: Diagram of the optical system. A biomolecule, represented as a sphere, is located inside a nanochannel, represented as a cylinder.

of the channel is several orders of magnitude larger than its axial dimensions, the wavelength of the incident light, and the portion of the channel being irradiated. With these assumptions in mind, the power P_c of the light scattered by a segment of length L of cross-sectional area A can be written in analogy to Eq. (2.1) as

$$P_c = cI_0 L\sigma_c \tag{2.5}$$

where the scattering cross section of the cylinder, σ_c is now per unit length, and depends on the polarizability of the channel α_c as follows:

$$\sigma_c = \frac{k_o^3}{4}|\alpha_c|^2 \tag{2.6}$$

We notice that the wavenumber in question is now k_o, which corresponds to the wave propagating though the optically transparent matrix of the nanofluidic chip but outside the nanochannel. Explicitly, for parallel and orthogonal polarization [29], the polarizability of the cylinder reads:

$$\left|\alpha_{c,\parallel}\right| = A\left|m^2 - 1\right| \tag{2.7}$$

$$\left|\alpha_{c,\perp}\right| = A\left|\frac{m^2-1}{m^2+1}\right| \tag{2.8}$$

or equivalently,

$$\left|\alpha_c\right| = A\left|\frac{m^2-1}{1+\nu(m^2-1)}\right| \tag{2.9}$$

in which we have defined $m = n/n_o$ as the ratio of the refractive index inside

the channel n and the refractive index of the surrounding medium n_o[4], and the parameter ν containing the polarization characteristics:

$$\nu = \begin{cases} 0; \text{ for parallel polarization} \\ \frac{1}{2}; \text{ for orthogonal polarization} \end{cases} \tag{2.10}$$

We have got an expression for the polarizability of a channel.

2.2.3 Biomolecule inside a nanochannel

Now it is time to combine what we have learned about light scattering by nanoparticles and nanochannels. We emphasize our interest in describing the situation of a nanofluidic channel containing a biomolecule, which is precisely the platform of NSM. It is possible to use Mie theory to analytically solve a sphere inside a cylinder[25–27], by expressing spherical harmonics in terms of cylindrical harmonics, which involves expansions into Hankel functions and Legendre polynomials [30]. However, a much simpler approach will be considered to facilitate more tangible conclusions later.

Over a century ago, Maxwell Garnett published a very elegant *homogenization theory* [31, 32], in which he extensively explored the modelling of dielectric properties of a complex colloidal medium as a single homogeneous *effective* medium, by considering the contributions of each species to the total dielectric function. This enables a direct relation between microscopic (polarizability) and macroscopic properties (dielectric function). A significant part of the mathematical framework of Garnett's idea relies on the Clausius-Mossotti relation (analog to the Lorentz-Lorenz equation), which describes the relationship between the dielectric function of a bulk material and the polarizability of its constituent molecules [33]. Essentially, it states that the overall dielectric function of a material is determined by the volumetric average polarizability of its individual molecules. We recall that the dielectric function is no more than the square of the refractive index, a very familiar quantity at this point, which will function as a mathematical and conceptual bridge to model the polarizability of the system composed by the nanochannel containing a small nanoparticle.

Let n_{eff} be the effective (total) refractive index of the system constituted by the channel containing the particle. Using the Clausius-Mossotti relation, this quantity can be written as follows [19, Sup. Info.]:

$$n_{\text{eff}}^2 \simeq n^2\left(1 + \frac{\rho_v \alpha_p}{1 - \nu \rho_v \alpha_p}\right) \tag{2.11}$$

in which the volumetric density is defined as $\rho_v = N_p/V_c$ and represents the number N_p of particles of polarizability α_p contained in a portion of the cylinder with volume $V_c = A \cdot L$. Now we consider α_T as the effective (total) polarizability of the system,

[4]The medium surrounding the nanochannel is precisely the material of the optically transparent matrix of the nanofluidic chip, which is usually SiO_2. The refractive index n in the interior corresponds to the fluid filling the channel, which is usually based on water.

which is essentially a modified version of the polarizability of the cylinder (Eq. (2.9)), in which we replace the refractive index inside the channel to take into account the effect of the scattering particle ($n \to n_{\text{eff}}$):

$$\alpha_T = A\left|\frac{\left(n_{\text{eff}}/n_o\right)^2 - 1}{1 + \nu\left(\left(n_{\text{eff}}/n_o\right)^2 - 1\right)}\right| \tag{2.12}$$

For nanoparticles with low polarizability (such as biomolecules) that are sparsely distributed in the nanochannel, and with a small difference in the refractive indices inside and outside of the channel ($n^2 - n_o^2 << 1$), Eq. (2.12) can be greatly reduced to

$$|\alpha_T| = |\alpha_c| + \frac{M(\nu)}{L}|\alpha_p| \tag{2.13}$$

In which we have defined the parametric function $M(\nu) = m^2 - 2\nu\left(m^2 - 1\right)$, so $M(\frac{1}{2}) = 1$ for orthogonal and $M(0) = n^2/n_o^2$ for parallel polarization. Equation (2.13) is a remarkable result; it states that the effective polarizability reduces to a linear combination of the polarizabilities of the channel and the particle.

Now, in complete analogy with Eq. (2.6), the effective (total) cross section σ_T of the system can be written as

$$\sigma_T = \frac{k_o^3}{4}|\alpha_T|^2 \tag{2.14}$$

and the total power of the light scattered by the system, P_T, can be written analogously to Eq. (2.5) as

$$P_T = cI_oL\sigma_T \tag{2.15}$$

Additionally, if we combine Eqs. (2.13) to (2.15), we can explicitly write the total scattering intensity power of the channel containing the particle:

$$\begin{aligned} P_T &= cI_oL\sigma_T \\ &= cI_oL\frac{k_o^3}{4}\left(|\alpha_c| + \frac{M(\nu)}{L}|\alpha_p|\right)^2 \end{aligned} \tag{2.16}$$

Expanding the right side of the last equation, we can express the squared of each polarizability in terms of the corresponding scattering power ($|\alpha_c^2| \propto P_c$ and $|\alpha_p^2| \propto P_p$) by the use of Eqs. (2.1), (2.2), (2.5) and (2.6). Moreover, if we assume that we measure the power of light scattered by a portion of the channel of length $L = 3\pi M(\nu)/2k_o$ (which is at least 4 times larger than the incident wavelength), we get an elegant equation that relates the scattering power of each contribution to the total power P_T:

$$P_T = P_c + P_p + 2\ \text{sign}(n - n_o)\sqrt{P_cP_p} \tag{2.17}$$

where we write sign(z) to denote the sign function, which extracts the sign of the argument $z = (n - n_o)$, that is, the sign of the difference between refractive indices

inside and outside the channel, which is contained in the expression for the polarizability of the channel. We recall that the scattering power, which is proportional to the cross section, scales with the squared volume of the corresponding scattering body. Since the volume of the channel is much larger than that of the particle, the scattering power of the channel is going to dominate strongly over that of the particle ($P_c >> P_p$), thus, the sole contribution of P_p will be practically negligible. Furthermore, in NSM experiments, the nanochannel is filled with water-based buffers ($n \simeq n_{\mathrm{H_2O}} = 1.33$) and the chip is fabricated in SiO_2 ($n_o = n_{\mathrm{SiO_2}} = 1.46$), so we can finally write the total power as

$$P_T - P_c \simeq -2\sqrt{P_c P_p} \propto |\alpha_c||\alpha_p| \tag{2.18}$$

Even though the scattering power of the particle alone is negligible compared to that of the channel, the difference in scattering power of the channel containing the particle and that of the empty channel still gives a non-negligible scattering contribution arising from the particle via a coupling [5] term of the form $|\alpha_c||\alpha_p|$. In the next section it will become evident why we are interested in the difference of light scattered by the nanochannel containing a particle and the nanochannel alone.

We summarize the learning outcomes of this section as follows:

1. The power of light scattered by some scattering body is proportional to the square of its volume.
2. Detecting the power of light scattered by an individual nanoparticle is challenging due to their small size.
3. The difference between the power of light scattered by a nanochannel containing a particle and that of the nanochannel alone results in a non-zero contribution, which arises from the interaction between the particle and the channel.

It is pertinent to emphasize that (2.18) is only valid for the particular case of $L = 3\pi M(\nu)/2k_o$, an assertion that holds only if we measure the total scattering power of a region of this specific length. However, since power is linearly proportional to intensity, this equation suggests that the difference in scattering intensity might also be connected to the polarizability of the particle. This idea will be explored in the next section, as it will be shown that the polarizability of a particle is in fact closely connected to the its molecular weight under appropriate conditions.

[5] This coupling term is often called *interference term*.

2.3 Determination of molecular weight and size

In the last section it was shown that the difference in measured scattering power between the channel-particle system and the empty channel is directly proportional to the polarizability of the particle (Equation (2.18)). However, such result is valid only for a specific length of the portion of the nanochannel. Below it is presented a similar relation which overcomes this limitation by defining a measurable quantity called integrated optical contrast (iOC).

2.3.1 Mass and integrated optical contrast (iOC)

We consider the intensity profile of the channel and the total system as $I_c(x) = cI_o\sigma_c$ and $I_t(x) = cI_o\sigma_t$, respectively, which is already integrated along the radial axis of the nanochannel. The integrated optical contrast (iOC) is defined as [19]:

$$iOC := \int_0^L \frac{I_T(x) - I_c(x)}{I_c(x)} dx \tag{2.19}$$

iOC represents an integrated value of the difference in intensities of scattered light normalized by the scattering intensity of the empty channel. With $\int_0^L I_T(x)dx = cLI_0\sigma_T$, the last integral results in

$$iOC = L\left(\frac{\sigma_T}{\sigma_c} - 1\right) \tag{2.20}$$

and substituting the scattering cross sections calculated in Eqs. (2.6), (2.13) and (2.14) we can express the iOC in terms of the relative polarizabilities of the channel and the particle

$$iOC = 2M(\nu_p)\frac{|\alpha_p|}{|\alpha_c|} + L\left(\frac{M(\nu_p)}{L}\frac{|\alpha_p|}{|\alpha_c|}\right)^2 \tag{2.21}$$

Since the polarizability is proportional to the volume, and the particle has a volume much smaller compared to that of the channel, the polarizability of the channel dominates strongly, so we can safely drop the squared term and reduce the last equation to

$$iOC \simeq \frac{2M(\nu_p)}{|\alpha_c|}|\alpha_p| \tag{2.22}$$

Substituting $|\alpha_c|$ (from Eqs. (2.6) and (2.9)) and recalling $M(\nu_p) = m^2 + 2\nu_p\left(1 - m^2\right)$ with $m = n^2/n_o^2$, we can simply write

$$iOC = \frac{\bar{n}(\nu_p)}{A}|\alpha_p| \tag{2.23}$$

with $\bar{n}(\nu_p)$ given by

$$\bar{n}(\nu_p) = \begin{cases} 2n^2/\left(n^2 - n_o^2\right); \text{ for orthogonal polarization } (\nu_p = 1/2) \\ \left(n^2 + n_o^2\right)/\left(n^2 - n_o^2\right); \text{ for parallel polarization } (\nu_p = 0) \end{cases} \tag{2.24}$$

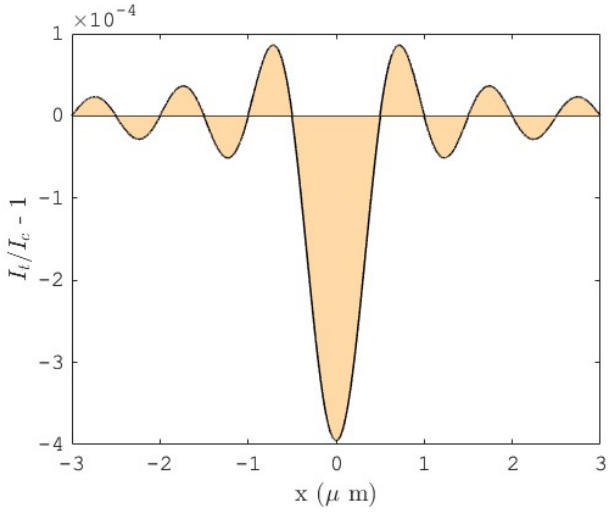


Figure 2.3: Illustration of (normalized) differential intensity along a nanochannel. Here, a particle is placed at $x = 0\mu m$, where the differential intensity has an minimum. iOC is the shaded area in yellow, defined as the area under the curve of the differential intensity. The optical signature of a biomolecule alone is hardly measurable, but can be inferred through the normalized difference in scattering intensities.

Note that both values of $\bar{n}(\nu_p)$ are negative since the refractive index of the optical matrix n_o is greater than the refractive index n of the fluid in the interior of the channel. For unpollarized light, we simply take the average of $\bar{n}(\nu_p)$ as

$$\langle \bar{n}(\nu_p) \rangle = \frac{\bar{n}(0) + \bar{n}(\frac{1}{2})}{2} = \frac{3n^2 + n_o^2}{2\Big(n^2 - n_o^2\Big)} \tag{2.25}$$

and substitute in Eq. (2.24)

$$iOC = \frac{\langle \bar{n}(\nu_p) \rangle}{A} |\alpha_p| = \frac{3n^2 + n_o^2}{2A\Big(n^2 - n_o^2\Big)} |\alpha_p| \tag{2.26}$$

Eq. (2.26) is a simple, linear, yet powerful relation that allows us to directly measure the polarizability of a single nanoparticle by measuring the normalized difference in intensity of the scattered light. In the following subsection, it will be shown that the polarizability of a particle is directly connected to its mass.

2.3.1.1 Polarizability, refractive index, and protein concentration: Connecting the dots

Now we intend to relate the polarizability of a nanoparticle with its mass. We will consider the case of proteins, since we are particularly interested in measuring biomolecules.

It is well known that the refractive index n_{prot} of a protein solution varies linearly with respect to concentration c_p [34–36], which can be written as

$$\frac{dn_{prot}}{dc_p} = 0.185 \text{ mL} \cdot g^{-1} = 0.307 \frac{\text{Å}^3}{\text{Da/mol}} \tag{2.27}$$

on the other hand, we can use the Maxwell Garnett´s mixing formula [33, Eq. 24] to write the refractive index of the solution as

$$n_{prot}^2 \simeq n^2 \left(1 + \frac{|\alpha_p|\rho_v}{1 - \frac{|\alpha_p|\rho_v}{3n^2}} \right) \tag{2.28}$$

in which n is the refractive index of the medium (usually water), $|\alpha_p|$ the polarizability of the protein and $\rho_v = N/V$ the number of molecules N in a volume V. The concentration of protein is linearly proportional to its molecular weight MW and the volume density it occupies, following $c_p = MW \cdot \rho_v / N_A$, with N_A the Avogadro's constant, so the differential operator d/dc_p becomes

$$\frac{d}{dc_p} := \frac{N_A}{MW} \frac{d}{d\rho_v}$$

which means that the rate of change in concentration is linearly proportional to the rate of change in density. Now, by assuming high dilution ($|\alpha_p|\rho_v << 1$) we can write the Taylor expansion for n_{prot} up to first term and substitute it in Eq. (2.27)

$$\begin{aligned} \frac{dn_{prot}}{dc_p} &= \frac{N_A}{MW} \frac{d}{d\rho_v} n_{prot} \\ &\simeq \frac{N_A}{MW} \frac{d}{d\rho_v} \left[n \Big(1 + \frac{|\alpha_p|\rho_v}{2} + \mathcal{O}\big(|\alpha_p|\rho_v\big)^2 + ... \Big) \right] \\ &\simeq \frac{|\alpha_p|}{MW} \frac{nN_A}{2} \end{aligned} \tag{2.29}$$

we group all the constant terms into a constant a', defined as

$$a' \equiv \frac{2}{N_A n} \frac{dn_{prot}}{dc_p} = 0.462 \frac{\text{Å}^3}{\text{Da}} \tag{2.30}$$

therefore we can corroborate that the polarizability is linearly proportional to the molecular weight:

$$|\alpha_p| = a' \cdot MW \tag{2.31}$$

lastly, we express the polarizability in terms of iOC by using Eq. (2.26) and write the molecular weight as follows

$$MW = iOC \frac{A}{\langle \bar{n}(\nu_p) \rangle a'} \tag{2.32}$$

We finally arrive at the following conclusion: The molecular weight of an individual biomolecule is proportional to the iOC [6], which can be measure with an

[6] Note that this relation is not only valid for proteins, but for any diluted species in which the rate of change in refractive index with respect to concentration is constant (Eq. (2.27)).

optical device. The factor $\langle \bar{n}(\nu_p) \rangle$ captures the optical properties of the media by taking into account the refractive indices, and A is the cross-sectional area of the nanochannel. a' captures the change rate in refractive index of a protein solution as its concentration varies, which is a well known quantity. All of these parameters can be either tuned or measured *a priori*. Now we have the framework to make single molecule measurements of mass.

The second stellar feature of NSM is the ability to estimate the size of a single particle, which is done by performing a statistical analysis of its trajectory. The next section will describe the relation between a trajectory and the size of a particle.

2.3.2 Particle size and diffusivity

Provided with an efficient particle tracking algorithm[19, 37], it is possible to extract trajectories of individual molecules from a kymograph (see Fig. 3.1a). Having a trajectory enables the possibility to perform statistical analysis and determine the rate at which a particle diffuses, which happens to be dependent on its size. This is done by the calculation of the diffusion coefficient or diffusivity, and its relation to the particle size via the Stokes-Einstein relation, which will be described below.

2.3.2.1 Brownian motion as a Wiener process

We consider a biomolecule diffusing in a nanochannel. The channel is filled with some fluid at room temperature, which at a molecular level, is a highly dynamic environment. The biomolecule is assumed to be larger than the constituent molecules of the fluid. We know that the temperature of the fluid is a measure of the average kinetic energy of its constituent molecules [38, p. 198], which on average move very fast and collide with each other at random[7]. Each of these elastic collisions transfer linear momentum to the biomolecule. After some finite time Δt, the biomolecule will be displaced by some distance Δx. Since this displacement is the result of the sum of a myriad of random collisions, by the central limit theorem we are safe to assume that Δx will follow a Gaussian distribution [40, p. 276]. Mathematically, this behaviour is well understood via a Wiener process, which is a fundamental, well-studied stochastic process [40], whose very useful properties will be discussed below.

We denote the position of a particle at a given time by $x(t)$. At each time increment dt, the particle exhibits a displacement dx given in general by:

$$dx = adt + bdW \tag{2.33}$$

in which a represents a drift rate, b^2 is the variance rate and dW is a differential increment of the Wiener process W, a random variable that follows a Gaussian distribution centered at zero with linearly increasing variance t, which is also

[7] Water molecules at room temperature collide with each other about 60 billion times every second [39, p. 107].

common to write as $\mathcal{N}(0, t)$.

Without loss of generality, we can redefine the net displacement as $dX := dx - adt$ and use the self-similarity properties of W [40, p. 525] to rewrite Equation (2.33) as

$$dX = b\sqrt{dt} \cdot \mathcal{N}(0, 1) \tag{2.34}$$

the mean squared displacement in infinitesimal form reads

$$\overline{dX^2} = b^2 dt \tag{2.35}$$

integrating with respect to time and setting $\overline{X(0)} = 0$

$$\overline{X^2(t)} = b^2 t \tag{2.36}$$

which is equal to the variance of position $Var\Big[X(t)\Big]$ [39, p. 116]

$$Var\Big[X(t)\Big] = \overline{X^2(t)} = b^2 t \tag{2.37}$$

In this formulation, the diffusion coefficient is equal to half the rate at which the variance of particle location changes over time [39, p. 115]:

$$D \equiv \frac{1}{2}\frac{d}{dt} Var\Big[X(t)\Big] \tag{2.38}$$

for discrete time series, $dt \to \Delta t$ and $dX \to \Delta X$, which allows us to write

$$D = \frac{1}{2}\frac{\overline{\Delta X^2}}{\Delta t} \tag{2.39}$$

Moreover, when analyzing real trajectories, we have to take into account that every frame is collected by a physical camera sensor, which captures light over some finite time interval and combines it into a single time frame, thus, smearing out the image of a moving object, which is known as motion blur. We use the covariance-based estimator to account for the motion blur and the localization error [41]:

$$D = \frac{1}{2}\frac{\overline{(\Delta X_n)^2}}{\Delta t} + \frac{\overline{\Delta X_n \Delta X_{n-1}}}{\Delta t} \tag{2.40}$$

In the last equation, the $\overline{(...)_n}$ symbol denotes averaging over all the time series points $n = t_n/\Delta t$. The second term is related to the covariance $\mathrm{Cov}(\Delta X_n, \Delta X_{n-1})$, which approaches to zero when motion blur is not present and the localization error is zero, thus coinciding exactly with the analytic result from Brownian dynamics (Eq. (2.39)). The last equation gives a simple way to compute the diffusion coefficient of single particles provided with their trajectory.

2.3.2.2 Diffusion coefficient and hydrodynamic radius

The diffusion coefficient D of the biomolecule is related to its hydrodynamic radius R_s according to the Stokes-Einstein relation

$$D = K\frac{k_B T}{6\pi\eta R_s} \tag{2.41}$$

where k_B is the Boltzmann constant, T the temperature, η the dynamic viscosity of the surrounding fluid and K is the hindrance factor [42], an empirical quantity that takes into account particle-wall hydrodynamic interactions and steric restrictions. The hydrodynamic radius R_s shall be interpreted as the radius of a hard neutral sphere which moves with diffusivity D.

The hindrance factor $K = K(\lambda)$ can be defined as [42]:

$$\begin{aligned} K(\lambda) = \Big(1 + 9\lambda \cdot \ln(\lambda/8) - 1.56\lambda + 0.53\lambda^2 + 1.92\lambda^3 - 2.81\lambda^4 + \\ 0.27\lambda^5 + 1.1\lambda^6 - 0.44\lambda^7\Big)\Big/\Big(1 - \lambda^2\Big) \end{aligned} \tag{2.42}$$

where $\lambda = R_s\sqrt{\pi/A}$, and A the cross sectional area of the channel. The hindrance factor monotonically decreases as λ increases (Fig. 2.4), due to λ approaching to 1 means that the particle's cross section is approaching that of the nanochannel, which results in an increasing hindrance in its freedom of movement, hence the name. For the same reason, the range of λ is from 0 to 1.

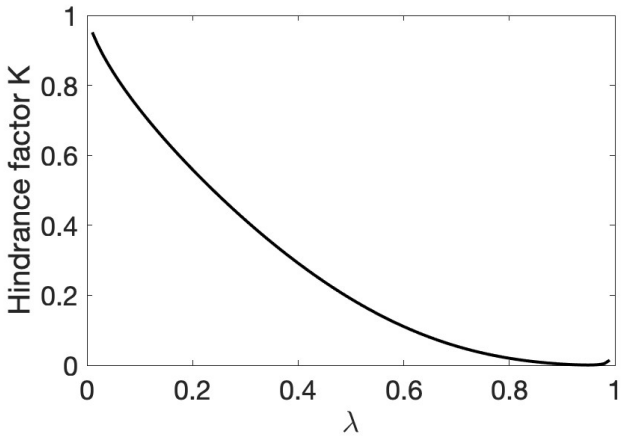


Figure 2.4: Hindrance factor as a function of $\lambda = R_s\sqrt{\pi/A}$. The monotonical decrease with λ corresponds to an increasing confinement effect; the motion of the particle becomes more constrained due to steric restrictions and particle-wall hydrodynamic interactions.

Having explored the theoretical aspects of Rayleigh scattering and the link between iOC and mass, along with the relationship between diffusion and particle hydrodynamic radius in some detail, we can now turn our attention to identifying common limitations of NSM and discussing how the theory can be expanded to address these challenges.

2.4 Common problems in NSM

The scope of this work is to find challenges in NSM which can be overcome with new data analysis algorithms *a posteriori*, that is, after the experiment has been performed. Two main difficulties were identified in NSM experiments: Biofouling and optical fringes. These phenomena are unrelated, thus, it is possible to treat them independently. Both cases are described below.

2.4.1 Biofouling

In experiments involving biomolecules flowing through micro and nanofluidics, particles may suffer an attractive interaction with the surface of the nanochannel (commonly made of SiO_2) thus getting immobilized for some time. This receives the name of biofouling [43][8]. This is particularly common when working with proteins. At a fundamental level, the explanation of this phenomenon is based in part on Van der Waals and electrostatic interactions between the particle and the inner walls of the channel. A characterization (and quantification) of such forces was firstly introduced by H. C. Hamaker in his acclaimed article from 1937 [44], in which he calculated Van der Waals interaction constants considering different geometries and dimensions. However, more recent and comprehensive studies have revealed that hydrophobic interactions play a major role in protein aggregation and adsorption [43, p. 68-76], being highly dependent on pH. When a protein adsorbs onto a surface, it unfolds and eventually undergoes denaturation. This can cause more proteins to aggregate near the inner wall of the nanochannel in a cascading effect.

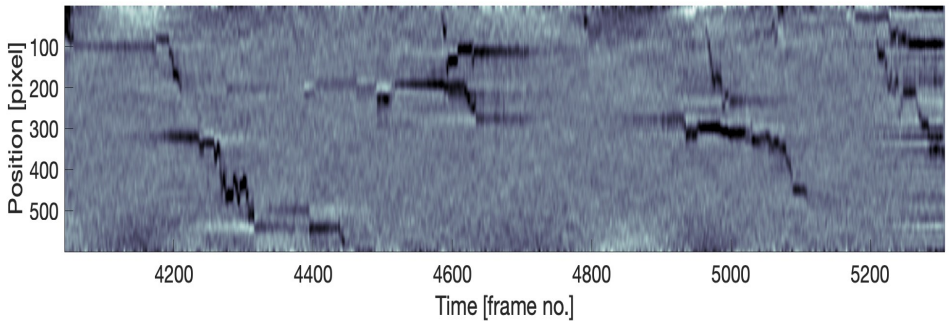


Figure 2.5: Kymograph from real data of a ferritin sample with evident biofouling. Each particle gets immobilized when diffusing through the channel, reason for which the trajectories appear horizontal (constant position over time).

Biofouling causes many problems in measurement instruments, such as reduced device sensitivity, decreased detection limits and a lower device lifetime [43].

[8] In fact, this is not restricted to micro and nanofluidics, but any time a surface is in contact with a fluid containing biological particles, regardless of scale or geometry, such as naval ships or drainage systems.

In NSM and other similar techniques, the particle size is calculated based on its trajectory following free diffusion. Since biofouling prevents a particle to undergo free diffusion, the particle size cannot be estimated. Therefore, it is imperative to reduce biofouling and prevent it from affecting the calculation of diffusivity.

It is crucial to understand the nature of the interactions that cause biofouling in order to apply the most effective solution. Nanofluidic chips are usually fabricated in SiO_2, which in water is negatively charged due to the formation of silanol ($Si-OH$) groups that tend to deprotonate at pH higher than 2 [45]. Nucleic acids in DNA and RNA are negatively charged [46], so they will be naturally repelled by the walls of the channel. However, proteins can acquire positive charge in solution, depending on the content of ionizable groups and their pKa values[9] [47]. This yields to an electrostatic attractive force, which partially explain why proteins are more prone to biofouling.

Various techniques can be applied to the nanofluidic chip to mitigate the previously mentioned effects. The simplest method is hydrodynamic actuation, which involves creating a pressure differential between the ends of the nanochannel to generate a net flow. Adding salts to the fluid can reduce electrostatic interactions through charge screening [48]. Additionally, surface modifications with PEG (polyethylene glycol) can decrease hydrophobic interactions between proteins and the inner wall of the channel. Nevertheless, modifying the inner surface of a nanochannel remains challenging, resulting in poor surface coverage and thus limited effectiveness of the surface passivation method. Therefore, at the level of data analysis, it is essential to take into account that trajectories might exhibit some degree of biofouling even when surface passivation has been implemented, for which as method is proposed in Section 3.2.

2.4.2 Optical fringes

While this work primarily addresses the Rayleigh approximation for light scattering because the particles measured in NSM are significantly smaller than the laser wavelength, there is a scenario where Mie scattering effects could be significant in experiments, specifically when the particle diameter approaches 100 nm and higher; the size of typical lipid nanoparticles (LNPs)[49].

When a particle of radius r gets irradiated with light of wavelength λ comparable to its radius ($r \simeq \lambda$), the assumption of Rayleigh scattering no longer holds. This means that the phase variation of the wave across the volume of the particle is no longer negligible, thus, the electric field is not uniform. In this case, the particle scatters light in a way described by the Mie scattering theory[50] [10]. In this regime, the resulting scattered wave is no longer symmetric with respect to the scattering

[9] pKa value is an acid dissociation constant which depends on structure and pH.

[10] The solution of Maxwell's equations for a spherical system involves a rather complicated series expansion into Hankel functions, which renders impossible to draw a simple, linear relation between iOC and molecular weight.

Figure 2.6: Optical fringes from a fragment of experimental kymograph of a measurement of LNPs, with diameters ranging from ∼50nm to ∼150 nm. Alongside the trajectories of big particles (thicker, darker trajectories), there is an adjacent, dimmer trajectory, which might be caused by Mie scattering since the size of the particle is comparable to the wavelength of the laser of visible light.

center (see Figs. 2.7 and 2.8), which shows as a characteristic optical fringe pattern in NSM. This pattern arises from the interference of the scattered light, and appears like ripples in intensity attenuating with the distance from the center of the particle (Fig. 2.6). These intensity oscillations are called optical fringes. Strictly speaking, they extend to infinity, but in reality only a few are detectable above the intrinsic noise level of the system. The distance between the optical fringes is constant during the experiment, because the optical system is in principle operating in stationary state and the interference pattern does not depend on the position of the particle [11]. The intensity of the optical fringes also decreases monotonically with distance.

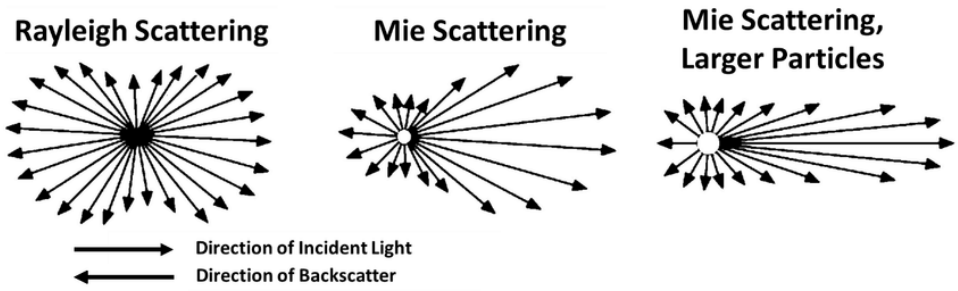


Figure 2.7: Schematics of the intensity and direction of propagation of scattered light. The incoming light comes from left to right. Note that for Rayleigh scattering, the intensity of scattered light is symmetrical with respect to the center of the particle. If the particle radius is comparable to the wavelength of incoming light, the intensity of backscattered light decreases and the intensity is no longer symmetrical with respect to the center of the particle. From Barnhart, Gunasekaran, 2020 [51].

[11] This holds as long as the molecular weight and radius of the particle is constant over time and space.

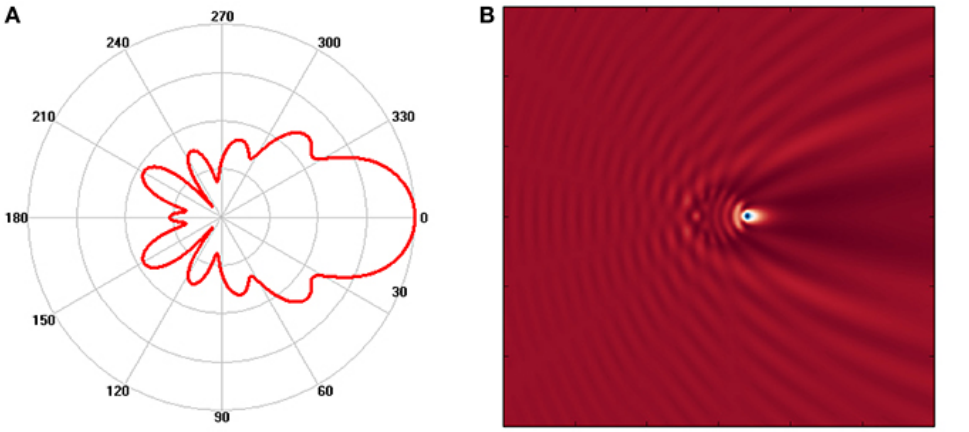


Figure 2.8: Scattered field produced by a PMMA sphere, with radius equal to the wavelength (1 μm), calculating using the T-matrix method (from Berisha, et. al., 2017 [52]). The scattering intensity has a preference along the direction of the incident beam (from left to right). The optical fringes emerge from the diffraction pattern of scattered light, which consequently is not symmetric. These fringes appear in the kymograph shown in Fig. 2.6, which might be mistakenly identified as additional particles.

Optical fringes represent a problem in NSM because the particle tracking algorithm may mistakenly identify them as an additional particle, moving essentially in the same manner as the real particle, but with a lower iOC. By Eqs. (2.32) and (2.41) the latter means that this "false" particle has the same hydrodynamic radius as the real particle but a lower molecular weight because the intensity of optical fringes decreases with distance. The detection of false positives will negatively affect the mass and size distribution of the total measurement, defeating the goal of NSM to be one of the most precise and accurate experimental single-molecule, label-free methods available.

3
Methods

The minimal physical components for NSM are a nanofluidic chip, an optical path powered by a laser of visible light, and a dark-field microscope equipped with a high speed camera (see Fig. 2.1). A detailed description of the components can be found in the main NSM article [19]. The camera is connected to a computer to enable real-time data acquisition using custom software in both Python and MATLAB®. The custom software used for data analysis is written in MATLAB and is described in detail in the Supplementary Information (SI) of the aforementioned article. The basic steps of the algorithm for data processing will be briefly outlined in the following subsections. Subsequently, the method proposal of this work will be presented in more detail, focusing on the mathematical derivation.

3.1 NSM algorithm overview

A CMOS camera captures the light scattered by the nanochannel containing the biomolecule(s) and collects it into a video file, which is no more than a 2D intensity map as a function of time. This file needs to be processed to visualize the biomolecule(s) and extract trajectories according to the following steps:

1) Integration along the nanochannel short axis. The raw signal is first integrated along the short axis of the nanochannel, resulting in an intensity map with only one spatial coordinate and a time coordinate.

2) Noise reduction. To reduce the effect of mechanical and optical fluctuations on the signal, Gaussian filtering is applied, which consists on the spatial convolution of the signal I_0 with a Gaussian kernel. This operation smoothes out the high frequency fluctuations, but also ensures the mean optical signal to remain unchanged, since the Gaussian kernel is normalized to preserve the norm. This procedure is particularly effective when the optical signal is also expected to exhibit a Gaussian profile.

3) Background estimation and first subtraction. The intensity profile of an empty channel $I^b(x,t)$ (considered background) is estimated by a spatio-temporal moving mean of the corrected total intensity profile $I^*(x,t)$. With this in mind, the normalized kymograph intensity profile $I^k(x,t)$ is calculated as:

$$I^k(x,t) = \frac{I^*(x,t) - I^b(x,t)}{I^b(x,t)} \tag{3.1}$$

which can be readily seen to be strongly related to the relations between iOC and molecular weight Eqs. (2.19) and (2.32). After this step, the optical signatures of particles diffusing through the nanochannel are revealed for the first time, yet are not ready for analysis.

4) Kymograph refinement. The optical signal of the kymograph, $I^k(x,t)$, is subjected to a thresholding process, in which a mask containing all the optical signatures is produced. An iterative process of noise reduction and background estimation and subtraction is performed, but this time the mask is used to exclude the particles' optical signatures to reduce the influence of denoising and refinement on the experimental intensity. Convergence is achieved when the variation of each matrix element of the intensity profile (absolute value) is less than some constant multiplied by the standard deviation of the optical profile. This constant is user-defined and sets the desired resolution of iOC, but in turn greatly affects the total computation time.

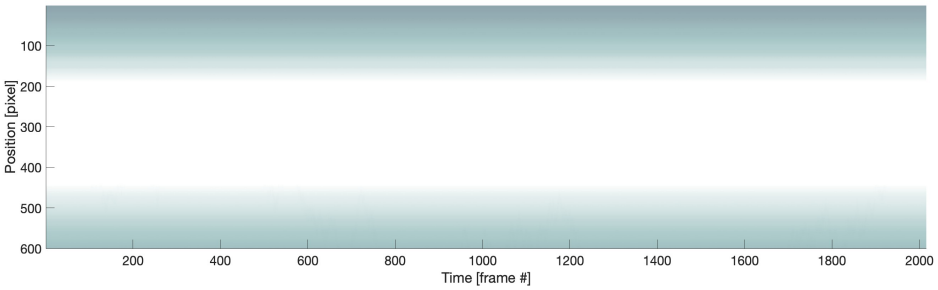
4) Particle tracking. The determination of the particle trajectories consists of several steps. First, feature points are identified, which correspond to significantly distinctive local minima of the intensity profile at each time frame n (by definition, the iOC is negative). There exists a maximum diffusion coefficient, denoted as D_{max}, that the particle tracking algorithm can detect. The standard deviation of the Gaussian probability distribution for the displacements, Δx corresponding to this diffusion coefficient is given by $\sqrt{2D_{max}\Delta t}$ (see Eq. (2.39)). To capture 99.73% of particles moving at this diffusion rate, a maximum displacement Δx_{max} is defined as three times this standard deviation. Thus, Δx_{max} serves as the upper limit for associating pairs of subsequent feature points. The algorithm first identifies pairs of feature points whose separation distance is less than or equal to Δx_{max}. This process is repeated iteratively, where each iteration looks for a connection between any two pairs of feature points within this threshold separation distance. The iterations continue until no further associations are possible within the limit Δx_{max}.

5) Data analysis. Diffusion coefficient can be extracted from a trajectory using an average of the mean square displacements per particle (MSD) as a function of time, accounting for localization and motion blur, as depicted in Eq. (2.40). Knowing the diffusion coefficient, the hydrodynamic radius R is calculated using Eq. (2.41), in which Newton-Rhapson method is applied to solve for R , since the hindrance factor is a complicated function of R. The time-dependent iOC is calculated by the integral of the normalized kymograph along position Eq. (2.19). The median over time is identified simply as iOC, and it is used to calculate the molecular weight of a biomolecule using Eq. (2.32).

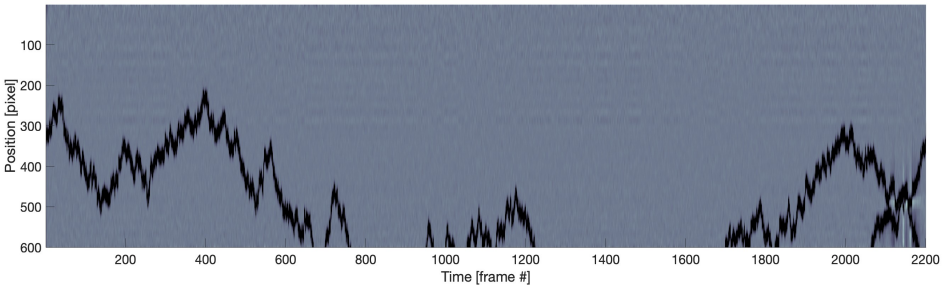
3.1.1 The kymograph

After completing data preprocessing (Steps 1-4), a kymograph is produced, which represents the intensity of light scattered by biomolecules within a nanochannel as a function of position and time. With a reliable particle tracking algorithm, the

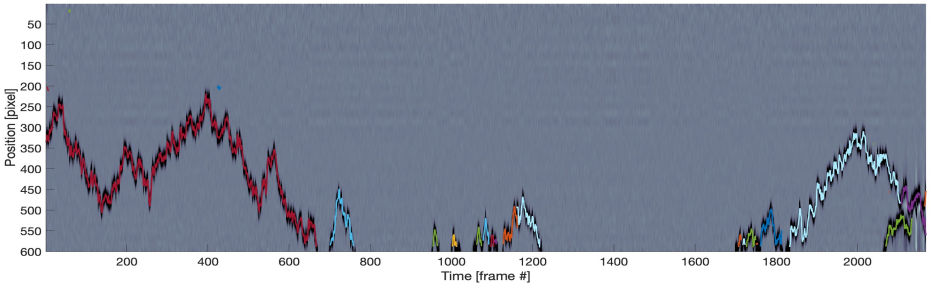
trajectories can be identified, enabling the calculation of the diffusion coefficient and the subsequent determination of the biomolecules' hydrodynamic radii.



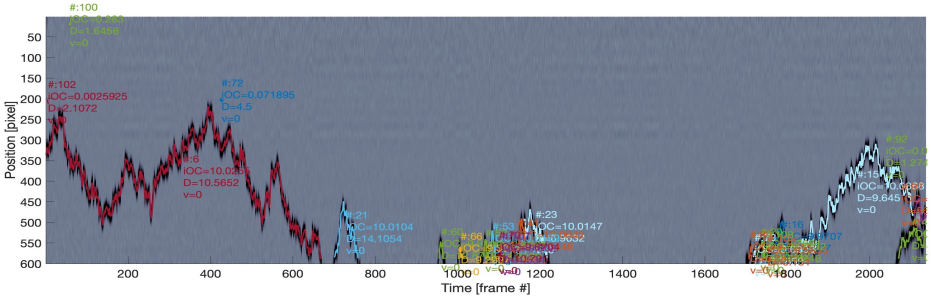
(a) Raw signal.



(b) Refined kymograph.



(c) Particle tracking algorithm.



(d) Analysis of trajectories to calculate D and iOC.

Figure 3.1: Stages of kymograph refinement and analysis. The vertical axis represents position inside the nanochannel, while the horizontal represents time. A particle tracking algorithm is applied to the kymograph to extract the trajectories of individual particles. Then, the diffusion coefficient D (Eq. (2.40)) and iOC are calculated for each individual trajectory.

3.2 Method proposal: Biofouling detection

Brownian motion is a stochastic process in which a particle can be displaced a random distance at any time interval. Under the hypothesis of free diffusion, these displacements follow a Gaussian distribution, and we can relate the variance of the trajectory to a diffusion coefficient, which yields to the hydrodynamic radius of the Brownian particle. If such a particle gets immobilized due to interactions with the channel, the hypothesis of free diffusion is no longer valid, thus, it is not possible to associate a diffusion coefficient to its trajectory. Therefore, we are interested in finding a method to determine when a trajectory satisfies the hypothesis of free diffusion, because this is the only way in which we can extract the particle size. In this section we will derive a method based on the statistical analysis of the instantaneous kinetic energy. We will see that the probability of a Brownian particle to retain a low kinetic energy decays exponentially with time. This is going to be implemented as an algorithm and evaluated with simulated and real trajectories.

3.2.1 Probability distribution of kinetic energy in Brownian motion

We recall the definition of instantaneous kinetic energy T as $\frac{1}{2}m\Big(\frac{dx(t)}{dt}\Big)^2$, with m the particle mass. Taking ΔX in terms of its probability distribution from Eq. (2.34), we can write the kinetic energy in discrete form as:

$$T = \frac{1}{2}m\Big(\frac{\Delta X}{\Delta t}\Big)^2 = \frac{1}{2}m\frac{b^2}{\Delta t}\Big[\mathcal{N}(0,1)\Big]^2 \tag{3.2}$$

the term $\Big[\mathcal{N}(0,1)\Big]^2$ refers to the squared of a standard Gaussian random variable, whose probability distribution is given by a standardized χ^2_k (Chi-squared) distribution with $k = 1$ degree of freedom [53, p. 940]:

$$\Big[\mathcal{N}(0,1)\Big]^2 = \chi^2_{k=1} = \frac{1}{\sqrt{2\pi}}\frac{e^{-\frac{y}{2}}}{\sqrt{y}} \tag{3.3}$$

In which y refers to the square of a dimensionless displacement which will be interpreted later. $\chi^2_{k=1}$ has convenient properties:

$$\begin{aligned} E\Big[\chi^2_{k=1}\Big] &= 1 \\ \mathrm{Var}[\chi^2_{k=1}] &= 2 \end{aligned} \tag{3.4}$$

and its cumulative distribution function CDF given by

$$CDF_{\chi^2_{k=1}}(y) = \frac{1}{\sqrt{\pi}}\int_0^y \frac{e^{-\frac{y'}{2}}}{\sqrt{y'}}dy' = \frac{\gamma(\frac{1}{2},\frac{y}{2})}{\Gamma(\frac{1}{2})} \tag{3.5}$$

where $\gamma(\frac{k}{2},\frac{y}{2})$ is known as the lower incomplete gamma function and $\Gamma(\frac{k}{2})$ the gamma function. In this case we only have $k = 1$ degrees of freedom.

The properties of the $\chi^2_{k=1}$ distribution (Eq. (3.4)) allow us to easily compute the mean kinetic energy $E\Big[T\Big]$:

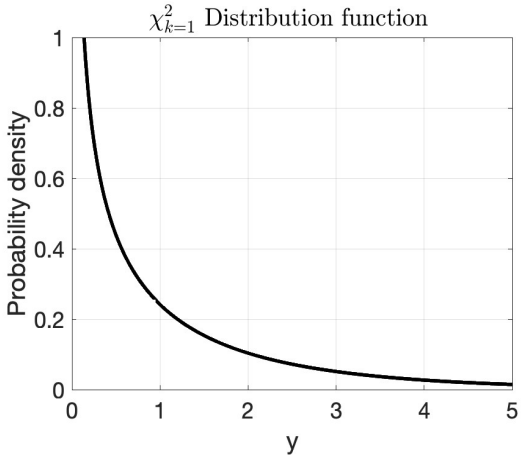
$$E\Big[T\Big] := E\Big[\frac{1}{2}m\frac{b^2}{\Delta t}\chi^2\Big] = \frac{1}{2}m\frac{b^2}{\Delta t}\cdot \cancelto{=1}{E\Big[\chi^2_{k=1}\Big]}$$
$$E\Big[T\Big] = \frac{1}{2}m\frac{b^2}{\Delta t} \tag{3.6}$$

then the kinetic energy normalized to the mean $\overline{T}_{norm}$ (dimensionless) is given by

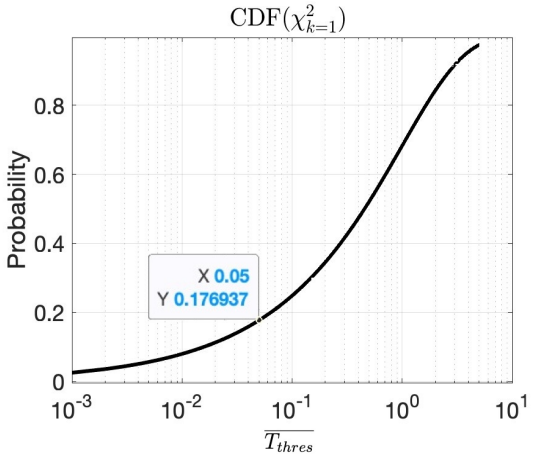
$$\overline{T}_{norm} := \frac{T}{E\Big[T\Big]} = \chi^2_{k=1} \tag{3.7}$$

This is an important result, which can be interpreted as follows:

For a particle undergoing free diffusion, the kinetic energy normalized to the mean is given by the standardized Chi-squared distribution with one degree of freedom.



(a) $\chi^2_{k=1}$ distribution function with k=1 degree of freedom, describing the kinetic energy normalized to the mean $\overline{T}_{norm}$.



(b) Cumulative distribution function of $\chi^2_{k=1}$. This is equal to the area under the curve of $\chi^2_{k=1}$. There is a probability of 17.7 % that $\overline{T}_{norm}$ adopts a value of 5 % or lower.

Figure 3.2: Chi-squared distribution and its corresponding cumulative distribution function, which describe the probability distribution of $\overline{T}_{norm}$.

In this way, we can interpret the cumulative distribution function evaluated at a certain threshold $\overline{T_{thres}}$ as the probability that, at any given time $t_n = n\Delta t$, the instantaneous kinetic energy normalized to the mean $\overline{T}_{norm}$ is at most $\overline{T_{thres}}$, which we write as

$$P(\overline{T}_{norm} \leq \overline{T_{thres}}) = CDF_{\chi^2_{k=1}}\Big(\overline{T_{thres}}\Big) \tag{3.8}$$

The quantity $\overline{T_{thres}}$ is the ratio between a certain kinetic energy threshold and the mean kinetic energy of the entire trajectory.

We can expand our reasoning further, by considering that each event of a discrete Wiener process W is statistically independent of the rest, thus, mathematically:

$$W_n - W_{n-1} = \mathcal{N}(0,1) \quad \forall \quad n > 1 \tag{3.9}$$

we also know that, given $a_1, a_2, ..., a_N$ statistically independent events, the joint probability of all of them occurring subsequently is

$$P(a_1|a_2|...|a_N) = \prod_{i=1}^{N} P(a_i) \tag{3.10}$$

combining equations Eqs. (3.5), (3.8) and (3.10), we can write the probability of a particle to retain a kinetic energy normalized to the mean $\overline{T}_{norm}$ remaining below or equal to $\overline{T_{thres}}$ for N steps as

$$P(\overline{T}_{norm} \leq \overline{T_{thres}}|\text{ N steps}) = \left(CDF_{\chi^2_{k=1}}\left(\overline{T_{thres}}\right)\right)^N \tag{3.11}$$

where the operation $(\cdot)^N$ denotes a simple exponentiation. We have derived Eq. (3.11) to allow the computation of the likelihood of a Brownian particle to stay with a low kinetic energy for a given extension of time. It can be seen from Fig. 3.2b that the probability of a particle to have an instantaneous $\overline{T}_{norm}$ lower or equal to 0.05 is 0.177. Therefore after, let's say, 5 steps, the probability will be

$$P(\overline{T}_{norm} \leq 0.05|\text{ 5 steps}) = (0.177)^5 = 1.173 \times 10^{-4} = 0.0173\% \tag{3.12}$$

evidently, this quantity decreases exponentially with the number of steps. It is valuable to note that this joint probability does not depend on the time step Δt, particle mass, global temperature, diffusion coefficient nor any of the physical quantities of the system; it is a general property of a particle following Brownian motion. Now we proceed to implement this result into an algorithm for the detection of biofouling.

3.2.2 Algorithm description

Particle trajectories have noise arising from the sequence of subroutines associated with particle tracking, which prevents a correct identification of biofouled segments. In this work, a simple noise reduction was performed using a moving average of the trajectory. It is recommend a sliding average time window of 5% of the total length of the trajectory. This ensures that the trajectory is minimally affected and the mean kinetic energy does not change significantly. An upper limit of 5 frames is imposed to the moving average time window in order to avoid over-smoothing of small features of the dynamics in longer trajectories. We also recommend the length of the sliding window to be odd, in order to correctly average to a central value, especially when the particle is immobilized, due to the noise is essentially random perturbations around an equilibrium, central position. We have previously derived Eq. (3.11), which gives the probability that a particle

retains at most a certain kinetic energy normalized to the mean $\overline{T_{thres}}$ for N time steps. In this work we considered the probability calculated in Equation (3.12) of $P(\overline{T}_{norm} \leq 0.05|\ 5 \text{ steps}) = 0.0173\%$ to be a significant indicator that such an event is unlikely to happen. However, these parameters shall be optimized and refined in future revisions. As discussed in 3.2 this probability threshold serves as a point of comparison; a way to evaluate the likelihood of a segment of a trajectory to take place.

Assuming a long trajectory with no biofouling ocurring, excluding all the detected segments with a normalized mean kinetic energy lower than 5% will only substract 0.0173% of the total trajectory, ensuring a low bias of the algorithm.

The main main steps of the algorithm are qualitatively described as follows:

1. Extract the trajectory (position and time) of a given particle by any particle tracking method.
2. Smooth the trajectory by moving average with a symmetric, odd sliding average window (denoising).
3. Calculate the square of the instantaneous velocity (linearly proportional to kinetic energy).
4. Calculate the mean square of the velocity.
5. Normalize the square of the velocity by the mean value (equivalent to kinetic energy normalized to the mean, since the constants cancel out).
6. Find all the segments longer than a certain amount of steps N with a normalized kinetic energy lower than a certain user-defined threshold $\overline{T_{thres}}$. These segments are designated as in "stuck mode", opposing to the "free diffusion mode".
7. Save the time and position segments corresponding to a stuck mode.
8. Proceed with the diffusivity calculation only considering segments corresponding to free diffusion.

This algorithm was written in MATLAB 2024, and it is found in Appendix A. At this level, T_{thres} and N are independent parameters, but the algorithm can be easily modified in order to set the probability of occurrence (see Equation (3.11)) constant and search for all the combinations of T_{thres} and N that satisfy this probability.

3.2.3 Simulation of trajectories with biofouling

The algorithms used to simulate kymographs are part of proprietary software, however, the algorithm to simulate the trajectories is based on standard Brownian motion, which is indeed a well known process. With the help of Equations (2.35) and (2.39), we can express the probability distribution of a small increment in position ΔX after a time interval Δt in terms of the diffusion coefficient D

$$\Delta X = \sqrt{2D\Delta t} \cdot \mathcal{N}(0,1) \tag{3.13}$$

which means that, at time t_n, we can express the position X_n as

$$X_n = X_{n-1} + \sqrt{2D\Delta t} \cdot \mathcal{N}(0,1) \tag{3.14}$$

This equation describes free diffusion in Brownian motion in discrete form. At each iteration, the current position is updated with the sum of the previous position and a random Gaussian variable with mean $\sqrt{2D\Delta t}$. Typical values of D for nanoparticles are in the range of $\mu m^2/s$ [54]. We set $\Delta t = 50$ ms to mimic the frame rate of the camera commonly used for NSM experiments at Envue Technologies.
Now we assume that biofouling occurs with low probability, that is, the total time of immobilization of a particle is about 1-5 % the total time of the experiment (50 s). This was performed by applying a "halt" order. Let x_{stuck} be the position at which a particle gets immobilized. Let t_{start} and t_{end} be the starting and ending times of biofouling, which are selected at random. Then we modify Equation (3.14) by neglecting the update of the position based on previous location, this is

$$X_n = X_{stuck} + \sqrt{2D\Delta t} \cdot \mathcal{N}(0,1); \text{ for } n = \frac{t_{start}}{\Delta t} \text{ to } \frac{t_{end}}{\Delta t} \tag{3.15}$$

This equation ensures that a particle stays at an equilibrium position, but takes into account thermal noise arising from the collisions with the molecules of the surrounding fluid. This equation was implemented to simulate biofouling, in which the amount of time spent in a fixed position was given by a weighted random, uniformly distributed variable, averaging out a total of 3-5% of the total time of the simulations. Simulations were carried out with 10000 time points, equivalent to 50 s in real time.

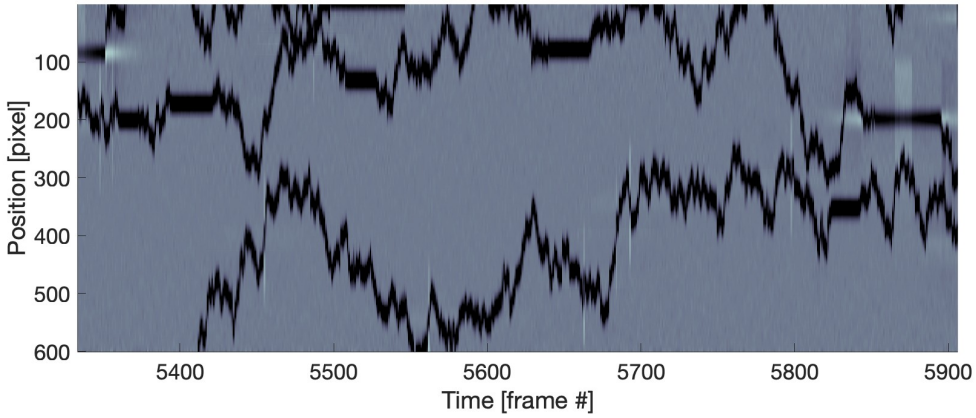


Figure 3.3: Simulated kymograph with diffusion coefficient $D = 30\mu m^2/s$ and $iOC = 10$ nm. Free diffusion is interrupted by biofouling, characterized for thick horizontal lines in the kymograph.

The iOC (see Equation (2.32)) in this work was set to 10 nm to optimize the signal to noise ratio and minimize any bias introduced by the algorithms responsible for kymograph refinement and particle tracking. If we consider a channel of area 50×30 nm^2, an iOC=10 nm is equivalent to about 3170 kDa or 3,170,000 g/mol for a globular protein, which is orders of magnitude higher than the largest known monomeric proteins [55]. However, at the level of simulations, iOC and diffusivity are independent parameters, so a high iOC does not affect the Brownian dynamics[1].

[1]In real experiments, iOC and diffusivity are actually interdependent because they are determined by mass and size, both being interrelated through density; an intensive, intrinsic quantity.

4
Results and discussion

A collection of kymographs were simulated using proprietary libraries from Špačková et. al.[19] and Envue Technologies. The pertinent parameters will be stated in each of the following subsections. This section outlines the results from the implementation of biofouling detection and its repercussions in the diffusion coefficient. We compare the change in the distribution of diffusivity before and after applying the algorithm. We relate them to distributions of hydrodynamic radius by setting a fixed channel area in all cases to facilitate comparison. The iOC was set high and constant on purpose, to reduce the number of variables to evaluate the performance of biofouling detection. Finally we apply the algorithm to experimental data of ferritin in PBS buffer, measured in January 2024 by Envue Technologies.

4.1 Qualitative proof of concept

The algorithm is able to detect segments of immobilized particles (Figure 4.2). The detection of such segments works best for trajectories longer than 40-50 data points, which is expected in probabilistic-statistical methods. Figure 4.1 illustrates different stages of the algorithm applied to one specific trajectory, as described in Algorithm description (Section 3.2.2). The first and last points of the biofouled segments are flagged with green and red points, respectively. Figure 4.2 shows the portion of the kymograph corresponding to the same trajectory shown in Figure 4.1. The biofouled segments of the trajectory in the kymograph are highlighted in white. The time scale is shown in dimensionless time steps, each step being equivalent to 1/200 seconds. Nevertheless, it was shown in Equation (3.11) that the kinetic energy normalized to the mean does not depend on any physical quantity, such as time step, diffusion coefficient or particle mass.

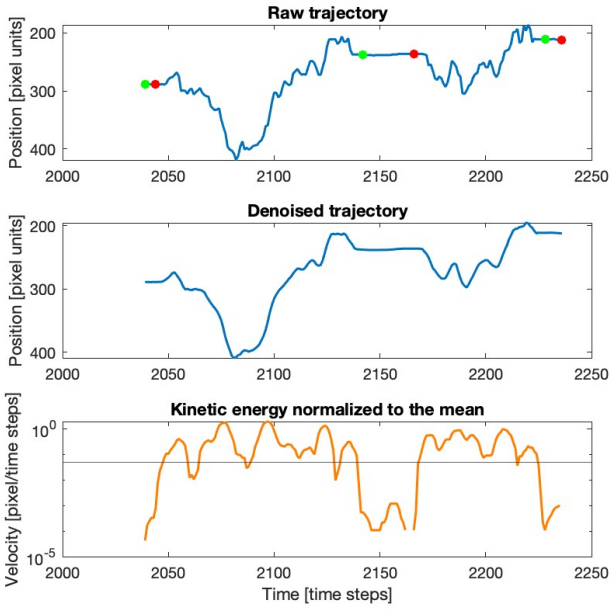
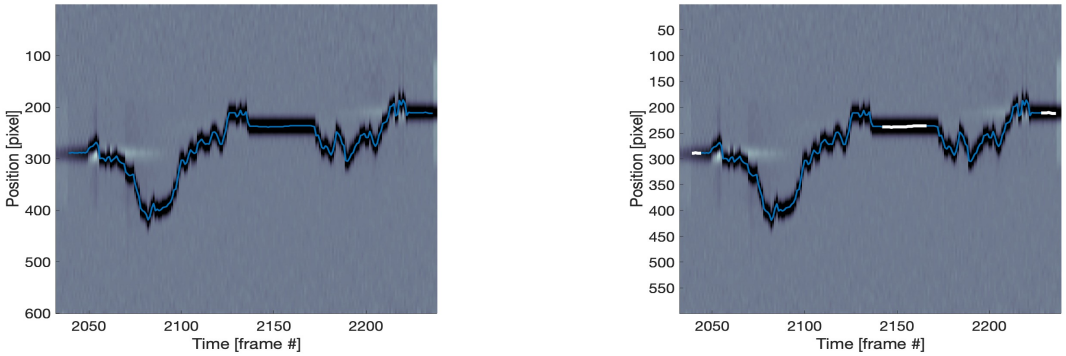


Figure 4.1: Example of biofouling detection. **Top**: Trajectory obtained from applying particle tracking algorithm to a simulated kymograph, with the starting and ending points of a biofouled segments highlighted in green and red, respectively. **Middle**: Trajectory denoised by moving average, with a sliding average time window of 2.5 % the total time. **Bottom**: Kinetic energy normalized to the mean in semilogarithmic scale. A horizontal line represents $\overline{T}_{thresh}$=0.05. Note that the segments in which $\overline{T}_{mean} < \overline{T}_{thresh}$ correspond to the segments delimited by the green and red dots in the top plot. The apparent discontinuity in kinetic energy arises from its value being very close to zero, which diverges to -Inf in logarithmic scale.



(a) Biofouling detection disabled.

(b) Biofouling detection enabled.

Figure 4.2: Simulated kymograph, corresponding to the same trajectory in Fig. 4.1. The biofouling detection algorithm correctly detects 3 segments (highlighted in white), as shown in Fig. 4.1.

4.2 Control test: Effect of biofouling detection in trajectories with no biofouling

As a control, biofouling detection was applied to simulated kymograph with no biofouling. Simulations of 30 kymographs, each containing more than a hundred trajectories, were performed with various diffusivity coefficients D = [10,20,30] $\mu m^2/s$ (10 kymographs each), with a fixed iOC = 10 nm. All trajectories follow standard Brownian motion with no biofouling.

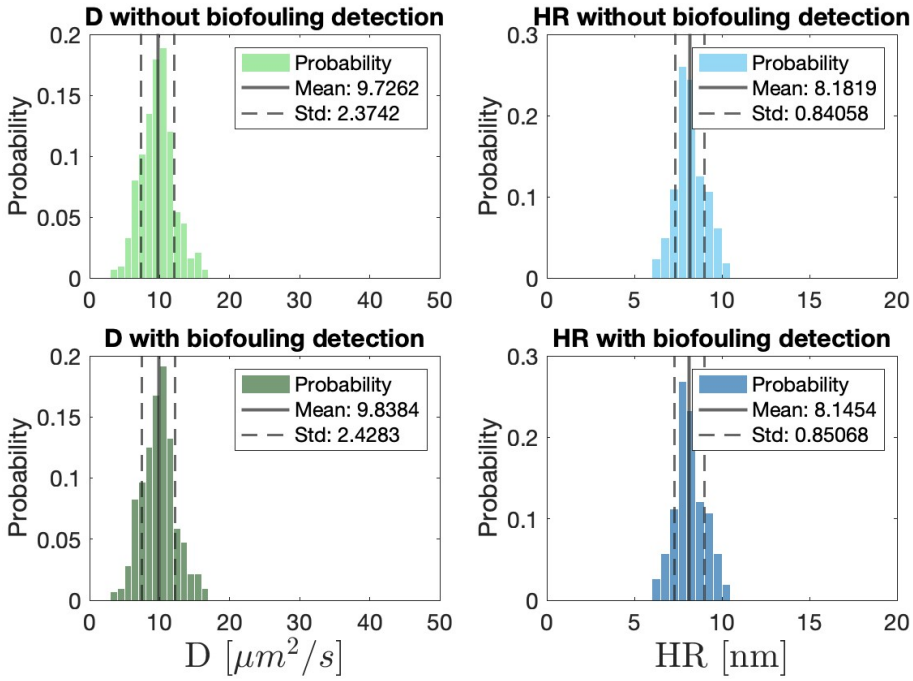


Figure 4.3: Comparison of distribution of diffusivities and corresponding hydrodynamic radius (HR) for a simulated diffusion coefficient $D = 10 \mu m^2/s$. The simulations were carried out without biofouling, so relatively small changes take place after applying biofouling detection. With a channel area of 50×30 nm^2, this diffusivity corresponds to an HR= 8.05 nm.

In Figure 4.3, the first row shows the distribution of diffusivity and hydrodynamic radius. The control test resulted in a mean diffusivity is 9.72 $\mu m^2/s$, differing by 2.3% from the target of D = 10 $\mu m^2/s$. This small deviation arises mostly from the fact that the dataset is finite, but for the purpose of this study is considered sufficiently valid. The second row shows these distributions after applying biofouling detection. Parameters for biofouling were set to $\overline{T}_{thresh} = 0.05$ and $N = 5$ (see Equation (3.12)). The mean diffusivity after biofouling detection resulted in 9.83 $\mu m^2/s$. Based on the control test, this constitutes a relative increase of 1.11%. Ideally, in the absence of biofouling, one would anticipate a null change. However, due to the probabilistic nature of the method, there exists a non-zero probability that the normalized kinetic energy will adopt low values for some amount of time. This increase represents the incidence of false positives at the level of diffusivity. Nevertheless,

since this alteration in mean diffusivity results in a bias in the mean hydrodynamic radius of approximately 0.35 Å, it is considered negligible, as it remains well below the detection limit of NSM [19].

4.3 Effect of biofouling in the calculation of diffusivity and hydrodynamic radius

Simulations of 30 kymographs, each containing around 100 trajectories on average, were performed with various diffusion coefficients D = [10,20,30] $\mu m^2/s$, with a fixed iOC = 10 nm. Biofouling was imposed at random, so a particle spends on average 3-5% of the total experiment time in a stuck position.

Figure 4.4 condenses the distribution of diffusivity for simulations with D =10 $\mu m^2/s$. Since only trajectories longer than 30 frames (150 ms) were considered, the dataset comprises a total of 837 trajectories. On top of the figure, the diffusivity was calculated using the whole trajectories.

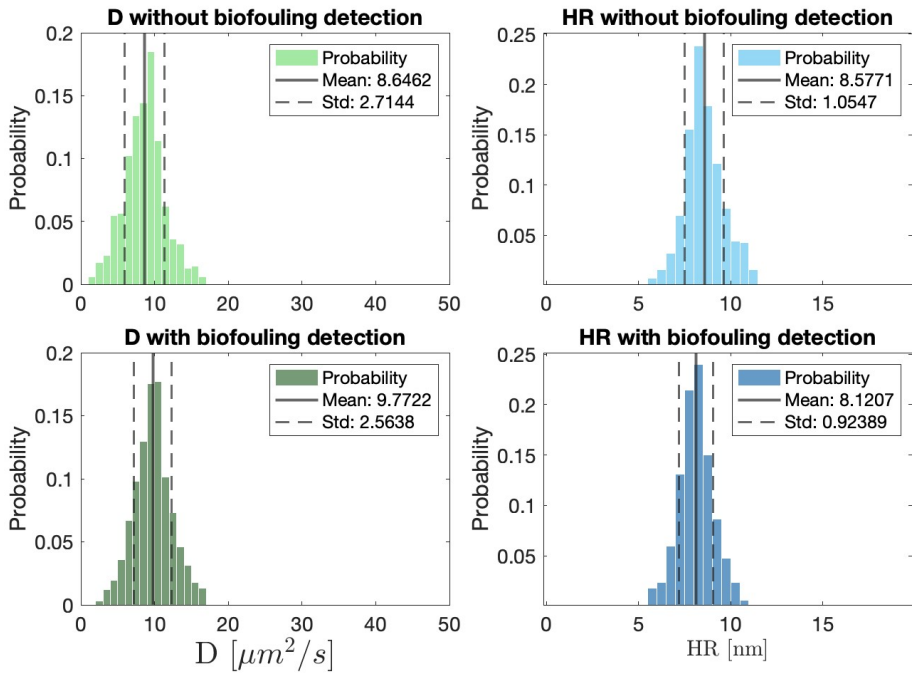


Figure 4.4: Changes before and after the detection of stuck trajectories. The target (simulated) diffusion coefficient was set to $D = 10\mu m^2/s$. A total of 837 trajectories longer than 30 data points we detected and analyzed. We exclude outliers lying more than 3 standard deviations from the mean. We notice that there is a significant improvement in the mean calculated diffusion coefficient (D), together with a reduction the standard deviation of both this quantity and the hydrodynamic radius (HR). HR was calculated assuming a channel area of 50×30 nm.

The distribution of diffusion coefficients for each trajectory is shown in Figure 4.4. The target diffusivity was set to $D = 10\mu m^2/s$. On top, the analysis was done with-

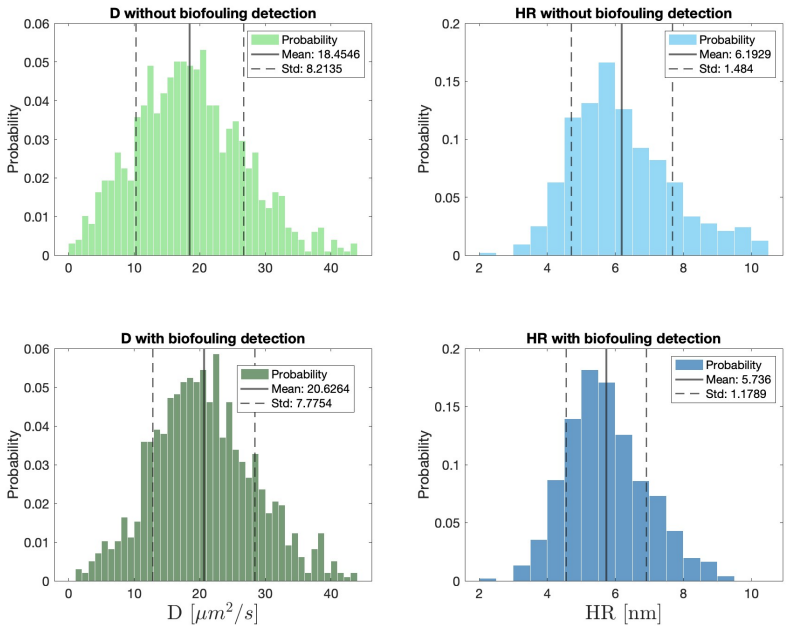


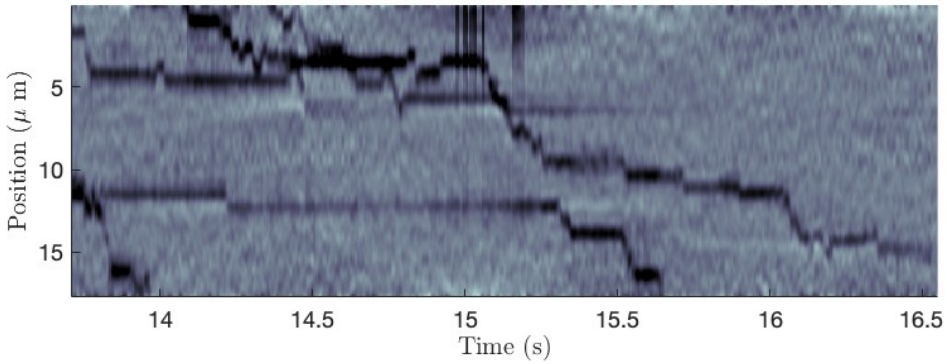
Figure 4.5: Changes before and after the detection of stuck trajectories. Simulated $D = 20\mu m^2/s$.

out excluding the biofouled segments from the trajectories, resulting in a decrease in estimated diffusivity.

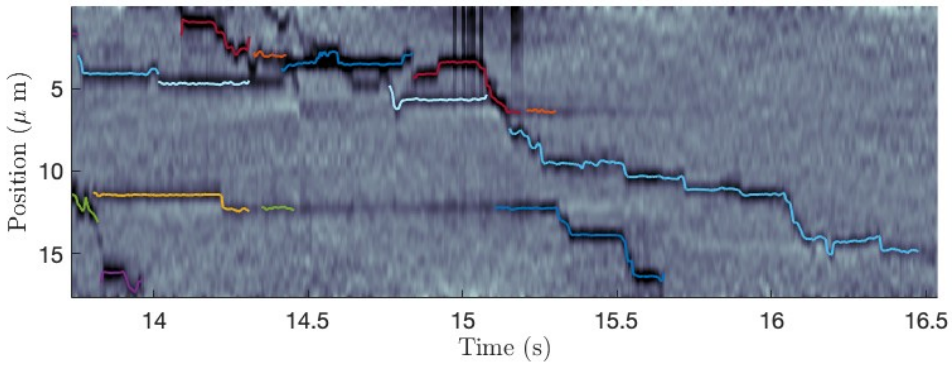
4.3.1 Real scenario: Ferritin oligomers

Ferritin is a self-assembling globular protein found in almost all living cells, acting as the main storage for iron [56], which makes it a valuable biomarker for detecting iron deficiency in blood. It has a typical radius of 6 nm [57, 58] and can store up to 4500 iron atoms, resulting in a mass ranging from about 400 kDa when empty to 750 kDa when fully loaded. Ferritin can form stable dimers, trimers and larger oligomers [57], and its cages have promising applications in drug delivery [59]. Thus, measuring the size of ferritin and the concentrations of its various oligomers is of significant interest in biomedicine.

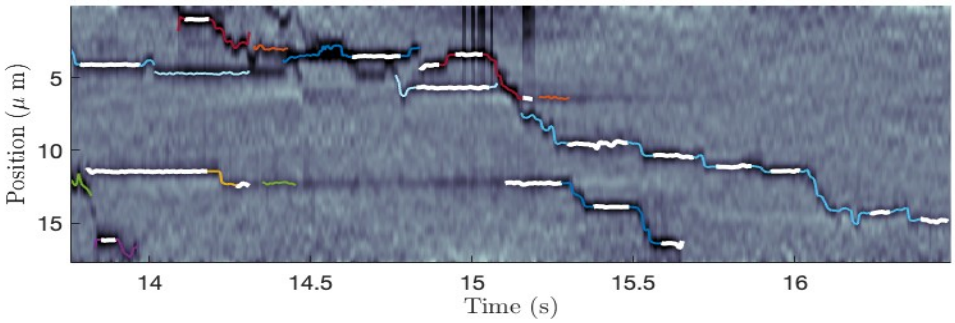
Biofouling detection was applied to experimental data of a series of ferritin samples in PBS, obtained in January 2024 by Envue Technologies. A total of 41 kymographs were analyzed. At the time of measurement, it was concluded that the surface passivation protocol (based on PEG) applied to the nanochannel was not sufficiently satisfactory, due to a heavy occurence of biofouling. Figure 4.6 illustrates the different stages of analysis for a portion of a single kymograph as an example. The detected biofouled segments are highlighted in white.



(a) Refined kymograph, before particle tracking.



(b) Refined kymograph, after particle tracking.



(c) Refined kymograph, after biofouling detection applied to trajectories (biofouled segments in white).

Figure 4.6: Different stages of a kymograph fragment of a ferritin sample presenting biofouling. The algorithm performs better for long trajectories.

Figure 4.6 shows a portion of a ferritin kymograph in which biofouling detection is applied. It is evident that the there is a significant presence of noise, since ferritin is a small particle, thus exhibiting low contrast and a lower signal-to-noise ratio. Despite of this, the particle tracking algorithm is able to identify complex trajectories, to which the biofouling detection algorithm is applied. The biofouled segments of the trajectory are colored in white. Performing a manual inspection, we realize that the algorithm can detect most of the biofouled segments, however, it performs better for long trajectories exhibiting both diffusion and biofouling,

because it is constructed under the assumption that free diffusion happens in the first place. For trajectories where biofouling is prevalent most of the time, a different method should be considered, possibly by identifying ranges of absolute kinetic energy per unit of mass that clearly indicate the particle is not undergoing free diffusion. These ranges of kinetic energy per unit of mass can be determined from the biofouled segments detected by the algorithm and implemented into a second iteration of biofouling detection.

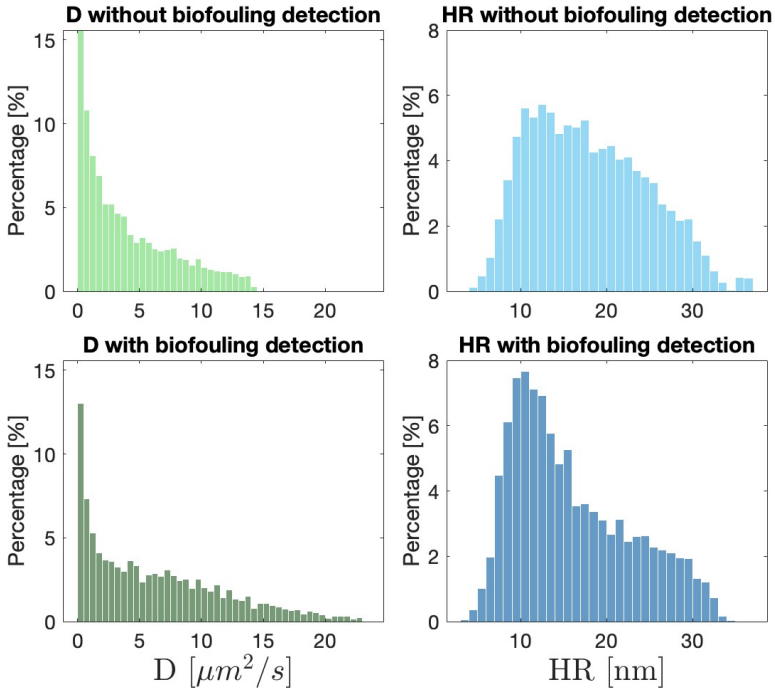


Figure 4.7: Diffusivity (D) and hydrodynamic radius (HR) distributions of all ferritin samples. After biofouling detection, a prominent presence of bodies appear in the range of HR $\sim$ 10-14 nm, suggesting a dominance of ferritin dimers.

The diffusivity and hydrodynamic radius distributions were calculated before and after biofouling detection (see Figure 4.7). In each instance, we discarded statistical outliers that were more than three standard deviations away from the median.

We observe that the distribution of diffusivity becomes broader when biofouling detection is applied, which is compatible with an increment in mean diffusivity, as previously noted in simulated data. However, there is an important occurrence of diffusivities approaching zero in both cases, which might be an indicative of completely biofouled trajectories not being excluded, as suggested in Figure 4.6. This offers additional motivation to expand the algorithms to include the case of fully biofouled trajectories.

Compared with the radius of a single ferritin unit (approximately 6 nm), the size distribution was considerably broad before biofouling detection, suggesting the

absence of a dominant contribution from any specific ferritin conformation. At the time of the experiment, it was determined that the sample was mainly composed of protein aggregates, with no clear prevalence of any particular ferritin configuration. After biofouling detection, a distinct peak appeared within the hydrodynamic radius range of about 10 to 14 nm. This suggests a significant contribution of ferritin dimers in the samples. However, these findings are provisional, since further optimizations in biofouling detection could shift this peak to the left, possibly indicating the presence of monomers.

We have learned that biofouling detection algorithms can help NSM extract physical information from experiments in which the antifouling protocols were not fully successful. Nevertheless, it is still imperative that particles diffuse through the nanochannel for most of the experiment time in order to apply the Stokes-Einstein relation, allowing the extraction of size distribution.

4.4 Limitations and future work

The biofouling detection algorithm presented in this work has been shown to be effective, although it has room for improvement. The method, which normalizes the instantaneous kinetic energy of a long trajectory primarily characterized by Brownian motion, fails in recognizing fully biofouled trajectories due to its reliance on self-comparison under the *a priori* assumption that only a small portion of the trajectory is biofouled. This is undesirable since the contributions of such trajectories greatly affect the diffusivity distribution. A plausible alternative to tackle this problem is described below. In addition, an algorithm for detecting optical fringes based on covariance analysis between trajectories is also outlined.

4.4.1 Detection of fully biofouled segments

It is reasonable to assume that a fully biofouled trajectory and a biofouled segment from a trajectory that exhibits both free diffusion and biofouling should share similar statistics, as they are governed by the same phenomena. From Equations (2.39) and (3.2), it follows that the mean (discrete) kinetic energy of a trajectory is proportional to the diffusion coefficient, since both are proportional to the mean squared displacement $\overline{(\Delta X)^2}$. By analyzing the fully biofouled segments of all the trajectories, we can get a distribution of biofouled mean square displacements. This distribution will serve as a reference to detect totally biofouled trajectories.

The new feature of fully biofouled detection is outlined in the following steps:

1. Utilize the preexisting biofouling detection algorithm to loop through all the N trajectories to detect biofouled segments.
2. Determine the mean squared displacement distribution of the m biofouled segments detected in the previous step, focusing on the average $\langle\overline{\Delta X_m^2}\rangle$ and standard deviation $\sigma_{(\Delta X_m)}$.
3. Loop over all the N trajectories again and compute the mean squared displacement $\overline{\Delta X_n^2}$ of the entire trajectory.

4. Apply the three-sigma rule: Any trajectory whose mean squared displacement $\overline{\Delta X_n^2}$ lies within a distance of $3\sigma(_{\Delta X_m})$ from $\langle\overline{\Delta X_m^2}\rangle$ will be considered a biofouled trajectory. Mark such trajectories.
5. Exclude the identified biofouled trajectories from the diffusivity calculations.

By implementing these steps, the algorithm can now take into account fully biofouled segments, which is expected to greatly improve the determination of the diffusivity distribution, hence the hydrodynamic radius.

4.4.2 Detection of optical fringes

In Section 2.4, it was described how the optical fringes can be identified by the particle tracking algorithm as an additional (false) particle, negatively impacting the reliability of the results after data have been processed. A key feature of optical fringes is that their separation distance remains constant over time in stationary state. This means that the real particle and its optical fringes describe trajectories that are *fully correlated* in the absence of noise. This assertion can be demonstrated by the use of the Pearson correlation coefficient.

Let $X = \{x_i\}$ and $Y = \{y_i\}$ be two discrete datasets of trajectories calculated by a particle tracking algorithm. The Pearson correlation coefficient r_{xy} is defined as

$$r_{x.y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{4.1}$$

In which $Cov(X,Y)$ is the covariance of X and Y, defined as:

$$Cov(X,Y) = \frac{\sum_{i=1}^{N}\Big(x_i - \langle x\rangle\Big)\Big(y_i - \langle y\rangle\Big)}{N} \tag{4.2}$$

and σ_Z being the standard deviation of a variable $Z = \{X,Y\}$, defined as usual:

$$\sigma_Z = \sqrt{\frac{\sum_{i=1}^{N}\Big(z_i - \langle z\rangle\Big)^2}{N}} \tag{4.3}$$

A correlation coefficient of $r_{x,y} = \pm 1$ indicates that the variables X and Y are fully correlated (positive sign) or anticorrelated (negative sign), that is, a linear change in X will cause a linear change in Y by the same magnitude [60]. Assuming that the dataset X describes a trajectory of a real particle and Y describes the trajectory of the maximum intensity of an optical fringe, detected as a false particle, we expect that at any i-th time frame, both positions will be related as:

$$x_i = y_i + d \tag{4.4}$$

in which d denotes the distance between the optical fringe and the real particle, assumed as constant. From Eqs. (4.1) to (4.3), it can be shown that the Pearson correlation coefficient between two variables is invariant under translations and

re-scaling. Therefore, we conclude the following:

In absence of noise, the trajectory of a particle and the trajectory that describes its optical fringe are fully correlated, with a Pearson correlation coefficient equal to 1.

In real NSM experiments, optical noise is unavoidable, so we should never expect a correlation coefficient of $r_{x,y} = 1$. However, it is well accepted that $r_{x,y} = 0.95$ is an indicator of high statistical significance [60]. With this in mind, it becomes possible to construct an algorithm for the detection of optical fringes as follows:

1. Loop over the N trajectories and calculate the Pearson correlation coefficient between particles m and n, with the condition $m > n$ to avoid double counting.
2. Detect all pairs of trajectories whose Pearson correlation coefficient $r_{m,n}$ is at least 0.95.
3. Compare the iOC of both trajectories and flag the one with the lowest absolute value.
4. Exclude all flagged trajectories from further data processing.

In step 3, we consider the trajectory with the highest absolute value of iOC, since an optical fringe always has a lower intensity.

This simple algorithm could be easily integrated into current data analysis tools and is expected to reduce false positive detection when measuring particles of diameters greater than 100 nm, as in the case of LNPs and exosomes [61].

5

Conclusion

NSM is an innovative single-molecule, label-free microscopy technique that integrates optics, electromagnetism, statistical physics, nanofabrication, surface chemistry, biophysics and computer science to determine the molecular weight of nanoparticles by measuring their polarizability, and their hydrodynamic radius by tracking their trajectory and calculating the corresponding diffusion coefficient. This technique utilizes a nanofluidic chip as the platform and induces light scattering in the Rayleigh regime. The optical signal is recorded and subsequently processed to extract single-molecule characteristics. This study aimed to explore the fundamental physical principles of NSM and identify limitations that could be addressed by enhancing data analysis methods. Biofouling emerged as a significant issue, a common challenge in micro and nanofluidics involving biomolecules, as it impedes the accurate determination of the diffusion coefficient, leading to a biased calculation of the hydrodynamic radius.

The study of Brownian motion properties has led to a method for determining the probability that a particle retains a certain kinetic energy over a given finite time interval. This principle was incorporated into an algorithm that analyzes each trajectory extracted from a kymograph, determining whether a drop in instantaneous kinetic energy (normalized to the mean) is due to the stochastic nature of free diffusion or is caused by biofouling.

As a control, biofouling detection was applied to simulated kymographs without ocurrence of biofouling. The algorithm was shown to introduce a low bias in the calculation of diffusivity, on the order of 1 to 2 % from the value calculated without biofouling detection. For simulations involving modest biofouling, in which each particle remained immobilized on average for 3 to 5% of the total duration of the experiment, biofouling detection brought the mean diffusivity back to the intervals given in the first analysis of kymographs without biofouling.

It is important to note that further evaluations of the biofouling detection algorithm need to be done to minimize false positive detection. An important step of the current algorithm is the denoising of trajectories by a uniform moving average time frame window. This moving window was taken to be at most 5% of the trajectory length. It is important to evaluate if the resulting data is still well described by Brownian motion, which can be done by an analysis of the velocity autocorrelation function.

As future research, an algorithm for identifying optical fringes based on correlation between detected trajectories could be easily implemented and tested, which could serve as an auxiliary tool to correct for optical artifacts when the particle size in question lies around the limit of validity of the Rayleigh scattering regime.

The studies presented in this work follow the notion that data analysis techniques grounded in fundamental physics and statistics can be applied to address some experimental challenges *a posteriori*, thereby lowering experiment costs and the required time to meet the desired accuracy and precision standards. In addition, a deeper understanding of the physical object of study is found in the journey.

Bibliography

1. Ball, P. Feynman's Fancy. *Royal Society of Chemistry.* https://www.rsc.org/images/Feynmans%20Fancy_tcm18-141620.pdf (Jan. 2009).
2. Gilbert, H. *Miniaturization* https://books.google.se/books?id=UxZTAAAAMAAJ (University of Michigan, 1961).
3. Toumey, C. Apostolic succession. *Caltech Magazine.* https://calteches.library.caltech.edu/4129/ (Jan. 2005).
4. Jones, R. Was Feynman the founder of nanotechnology? *Soft Machines.* http://www.softmachines.org/wordpress/?p=246 (Aug. 2006).
5. *Press release: The 2023 Nobel Prize in Physiology or Medicine* Oct. 2023. https://www.nobelprize.org/prizes/medicine/2023/press-release/.
6. Ishida, T., Murayama, T., Taketoshi, A. & Haruta, M. Importance of Size and Contact Structure of Gold Nanoparticles for the Genesis of Unique Catalytic Processes. *Chemical Reviews* **120,** 464–525. ISSN: 1520-6890. http://dx.doi.org/10.1021/acs.chemrev.9b00551 (Dec. 2019).
7. Rosário, F. *et al.* Impact of Particle Size on Toxicity, Tissue Distribution and Excretion Kinetics of Subchronic Intratracheal Instilled Silver Nanoparticles in Mice. *Toxics* **10,** 260. ISSN: 2305-6304. http://dx.doi.org/10.3390/toxics10050260 (May 2022).
8. Miao, C. *et al.* The size-dependent in vivo toxicity of amorphous silica nanoparticles: A systematic review. *Ecotoxicology and Environmental Safety* **271,** 115910. ISSN: 0147-6513. http://dx.doi.org/10.1016/j.ecoenv.2023.115910 (Feb. 2024).
9. Mollarasouli, F. & Bahrani, S. in *The Detection of Biomarkers* 23–69 (Elsevier, 2022). ISBN: 9780128228593. http://dx.doi.org/10.1016/B978-0-12-822859-3.00012-2.
10. Lemaire, B. *et al.* Ferritin: A Biomarker Requiring Caution in Clinical Decision. *Diagnostics* **14,** 386. ISSN: 2075-4418. http://dx.doi.org/10.3390/diagnostics14040386 (Feb. 2024).
11. Cunha, W. R. *et al.* High molecular weight adiponectin as a biomarker of hypertension in children and adolescents with obesity. *European Journal of Pediatrics* **182,** 2925–2931. ISSN: 1432-1076. http://dx.doi.org/10.1007/s00431-023-04962-w (Apr. 2023).
12. Diep Nguyen, T. Adiponectin: Role in physiology and pathophysiology. *International Journal of Preventive Medicine* **11,** 136. ISSN: 2008-7802. http://dx.doi.org/10.4103/ijpvm.IJPVM_193_20 (2020).
13. Pray, L. A. Discovery of DNA Structure and Function: Watson and Crick. *Nature Education* (2008).

14. Aug. 2024. https://www.sciencehistory.org/education/scientific-biographies/james-watson-francis-crick-maurice-wilkins-and-rosalind-franklin/#:~:text=These%20four%20scientists%E2%80%94Crick%2C%20Franklin,first%20broached%20by%20Maurice%20Wilkins..
15. Palade, G. E. A SMALL PARTICULATE COMPONENT OF THE CYTOPLASM. *The Journal of Cell Biology* **1,** 59–68. ISSN: 0021-9525. http://dx.doi.org/10.1083/jcb.1.1.59 (Jan. 1955).
16. *The Nobel Prize in Physiology or Medicine 1962* 2024. https://www.nobelprize.org/prizes/medicine/1962/summary/.
17. *The Nobel Prize in Physiology or Medicine 1974* 2024. https://www.nobelprize.org/prizes/medicine/1974/summary/.
18. Kamanzi, A. *et al.* Simultaneous, Single-Particle Measurements of Size and Loading Give Insights into the Structure of Drug-Delivery Nanoparticles. *ACS Nano* **15,** 19244–19255. ISSN: 1936-086X. http://dx.doi.org/10.1021/acsnano.1c04862 (Nov. 2021).
19. Špačková, B. *et al.* Label-free nanofluidic scattering microscopy of size and mass of single diffusing molecules and nanoparticles. *Nature Methods* **19,** 751–758. ISSN: 1548-7105. http://dx.doi.org/10.1038/s41592-022-01491-6 (May 2022).
20. Zhang, Y. *et al.* Fabrication of Nanochannels. *Materials* **8,** 6277–6308. ISSN: 1996-1944. http://dx.doi.org/10.3390/ma8095304 (Sept. 2015).
21. Cho, Y., Lee, S., Fujii, T. & Kim, B. Fabrication of silicon dioxide nanochannel arrays without nanolithography for manipulation of DNA molecule. *Microelectronic Engineering* **85,** 1275–1277. ISSN: 0167-9317. http://dx.doi.org/10.1016/j.mee.2008.03.001 (May 2008).
22. Reisner, W. *et al.* Nanoconfinement-Enhanced Conformational Response of Single DNA Molecules to Changes in Ionic Environment. *Physical Review Letters* **99.** ISSN: 1079-7114. http://dx.doi.org/10.1103/PhysRevLett.99.058302 (Aug. 2007).
23. Brooke, D., Movahed, N. & Bothner, B. Universal buffers for use in biochemistry and biophysical experiments. *AIMS Biophysics* **2,** 336–342. ISSN: 2377-9098. http://dx.doi.org/10.3934/biophy.2015.3.336 (2015).
24. Janeiro, R., Flores, R. & Viegas, J. Refractive index of phosphate-buffered saline in the telecom infrared C+L bands. *OSA Continuum* **4,** 3039. ISSN: 2578-7519. http://dx.doi.org/10.1364/OSAC.434864 (Nov. 2021).
25. McDonald, K. T. Scattering of a Plane Electromagnetic Wave by a Small Conducting Cylinder. *Joseph Henry Laboratories, Princeton University.* http://kirkmcd.princeton.edu/examples/small_cylinder.pdf (Oct. 2011).
26. Bohren, C. F. & Huffman, D. R. *Absorption and Scattering of Light by Small Particles* ISBN: 9783527618156. http://dx.doi.org/10.1002/9783527618156 (Wiley, Apr. 1998).
27. Hahn, D. W. *Light Scattering Theory* July 2009. http://plaza.ufl.edu/dwhahn/Rayleigh%20and%20Mie%20Light%20Scattering.pdf.
28. Kerker, M. Lorenz–Mie Scattering by Spheres: Some Newly Recognized Phenomena. *Aerosol Science and Technology* **1,** 275–291. ISSN: 1521-7388. http://dx.doi.org/10.1080/02786828208958594 (July 1982).

29. Altenburger, B. *et al.* Label-Free Imaging of Catalytic H2O2 Decomposition on Single Colloidal Pt Nanoparticles Using Nanofluidic Scattering Microscopy. *ACS Nano* **17,** 21030–21043. ISSN: 1936-086X. http://dx.doi.org/10.1021/acsnano.3c03977 (Oct. 2023).
30. Mangini, F. & Tedeschi, N. Scattering of an electromagnetic plane wave by a sphere embedded in a cylinder. *Journal of the Optical Society of America A* **34,** 760. ISSN: 1520-8532. http://dx.doi.org/10.1364/JOSAA.34.000760 (Apr. 2017).
31. Colours in metal glasses and in metallic films. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **203,** 385–420. ISSN: 2053-9258. http://dx.doi.org/10.1098/rsta.1904.0024 (Jan. 1904).
32. Garnett, M. Colours in metal glasses, in metallic films, and in metallic solutions. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **205,** 237–288 (Jan. 1906).
33. Markel, V. A. Introduction to the Maxwell Garnett approximation: tutorial. *Journal of the Optical Society of America A* **33,** 1244. ISSN: 1520-8532. http://dx.doi.org/10.1364/JOSAA.33.001244 (June 2016).
34. Khago, D., Bierma, J. C., Roskamp, K. W., Kozlyuk, N. & Martin, R. W. Protein refractive index increment is determined by conformation as well as composition. *Journal of Physics: Condensed Matter* **30,** 435101. ISSN: 1361-648X. http://dx.doi.org/10.1088/1361-648X/aae000 (Oct. 2018).
35. Barer, R. & Joseph, S. Refractometry of living cells part I. basic principles. *Journal of Cell Science* **S3-95,** 399–423. ISSN: 1477-9137. http://dx.doi.org/10.1242/jcs.s3-95.32.399 (Dec. 1954).
36. Hand, D. B. The refractivity of protein solutions. *Journal of Biological Chemistry* **108,** 703–707. ISSN: 0021-9258. http://dx.doi.org/10.1016/S0021-9258(18)75261-7 (Mar. 1935).
37. Sbalzarini, I. & Koumoutsakos, P. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology* **151,** 182–195. ISSN: 1047-8477. http://dx.doi.org/10.1016/j.jsb.2005.06.002 (Aug. 2005).
38. Greiner, W., Neise, L. & Stocker, H. *Thermodynamics and statistical mechanics* 1st ed. en (Springer, New York, NY, July 2000).
39. Denny, M. W. & Gaines, S. *Chance in biology* en (Princeton University Press, Princeton, NJ, Nov. 2000).
40. Olofsson, P. & Andersson, M. *Probability, statistics, and stochastic processes* 2nd ed. en (Wiley-Blackwell, Hoboken, NJ, May 2012).
41. Vestergaard, C. L., Blainey, P. C. & Flyvbjerg, H. Optimal estimation of diffusion coefficients from single-particle trajectories. *Physical Review E* **89** (Feb. 2014).
42. Dechadilok, P. & Deen, W. M. Hindrance Factors for Diffusion and Convection in Pores. *Industrial and Engineering Chemistry Research* **45,** 6953–6959. ISSN: 1520-5045. http://dx.doi.org/10.1021/ie051387n (Mar. 2006).

43. Yoon, J.-Y. & Garrell, R. L. *Biomolecular Adsorption in Microfluidics* (ed Li, D.) 68–76. ISBN: 978-0-387-48998-8. https://doi.org/10.1007/978-0-387-48998-8_87 (Springer US, Boston, MA, 2008).
44. Hamaker, H. The London—van der Waals attraction between spherical particles. *Physica* **4,** 1058–1072. ISSN: 0031-8914. http://dx.doi.org/10.1016/S0031-8914(37)80203-7 (Oct. 1937).
45. Lowe, B. M., Skylaris, C.-K. & Green, N. G. Acid-base dissociation mechanisms and energetics at the silica–water interface: An activationless process. *Journal of Colloid and Interface Science* **451,** 231–244. ISSN: 0021-9797. http://dx.doi.org/10.1016/j.jcis.2015.01.094 (Aug. 2015).
46. Lipfert, J., Doniach, S., Das, R. & Herschlag, D. Understanding Nucleic Acid–Ion Interactions. *Annual Review of Biochemistry* **83,** 813–841. ISSN: 1545-4509. http://dx.doi.org/10.1146/annurev-biochem-060409-092720 (June 2014).
47. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Science* **10,** 1206–1215. ISSN: 1469-896X. http://dx.doi.org/10.1110/ps.440101 (June 2001).
48. Perez-Jimenez, R., Godoy-Ruiz, R., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. The Efficiency of Different Salts to Screen Charge Interactions in Proteins: A Hofmeister Effect? *Biophysical Journal* **86,** 2414–2429. ISSN: 0006-3495. http://dx.doi.org/10.1016/S0006-3495(04)74298-8 (Apr. 2004).
49. Li, S. *et al.* Payload distribution and capacity of mRNA lipid nanoparticles. *Nature Communications* **13.** ISSN: 2041-1723. http://dx.doi.org/10.1038/s41467-022-33157-4 (Sept. 2022).
50. Lockwood, D. J. in *Encyclopedia of Color Science and Technology* (ed Luo, M. R.) (Springer, New York, NY, 2016). https://doi.org/10.1007/978-1-4419-8071-7_218.
51. Barnhart, S. & Sidaard Gunasekaran. Design and Development of a Coherent Detection Rayleigh Doppler Lidar System for Use as an Alternative Velocimetry Technique in Wind Tunnels. en. http://rgdoi.net/10.13140/RG.2.2.32670.15688 (2020).
52. Berisha, S., van Dijk, T., Bhargava, R., Carney, P. S. & Mayerich, D. BIM-Sim: Interactive Simulation of Broadband Imaging Using Mie Theory. *Frontiers in Physics* **5.** ISSN: 2296-424X. http://dx.doi.org/10.3389/fphy.2017.00005 (Feb. 2017).
53. *Handbook of mathematical functions* (eds Abramowitz, M. & Stegun, I. A.) (Dover Publications, Mineola, NY, June 1965).
54. Giorgi, F. *et al.* The influence of inter-particle forces on diffusion at the nanoscale. *Scientific Reports* **9.** ISSN: 2045-2322. http://dx.doi.org/10.1038/s41598-019-48754-5 (Sept. 2019).
55. Reisner, A., Rowe, J. & Macindoe, H. M. The largest known monomeric globular proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **188,** 196–206. ISSN: 0005-2795. http://dx.doi.org/10.1016/0005-2795(69)90066-X (Sept. 1969).

56. Mackenzie, E. L., Iwasaki, K. & Tsuji, Y. Intracellular Iron Transport and Storage: From Molecular Mechanisms to Health Implications. *Antioxidants and Redox Signaling* **10,** 997–1030. ISSN: 1557-7716. http://dx.doi.org/10.1089/ars.2007.1893 (June 2008).
57. Niitsu, Y. & Listowsky, I. Mechanisms for the formation of ferritin oligomers. *Biochemistry* **12,** 4690–4695. ISSN: 1520-4995. http://dx.doi.org/10.1021/bi00747a023 (Nov. 1973).
58. Domínguez, D. J. M. *Ferritin structure and its biomedical implications* June 2016. https://web.archive.org/web/20160827215820/http://www.ugr.es/~josema/Structure.html.
59. Zhu, Y. *et al.* Ferritin-based nanomedicine for disease treatment. *Medical Review* **3,** 49–74. ISSN: 2749-9642. http://dx.doi.org/10.1515/mr-2023-0001 (Feb. 2023).
60. Schober, P., Boer, C. & Schwarte, L. A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia* **126,** 1763–1768. ISSN: 0003-2999. http://dx.doi.org/10.1213/ANE.0000000000002864 (May 2018).
61. Rana, R., Kumari, B., Kumari, J. & Ganguly, N. K. Glioblastoma diagnostics and prognostic biomarkers: Current status in medicine and exosome derivation. *Current Medicine Research and Practice* **9,** 65–73. ISSN: 2352-0817. http://dx.doi.org/10.1016/j.cmrp.2019.03.001 (Mar. 2019).

A

Appendix 1: Algorithm for the detection of biofouling

The function *getStuckTrajectories* was written in MATLAB® 2023. It detects a drop in kinetic energy normalized to the mean if it is statistically significant (as discussed in Section 2.3.2.1).

```
    function PARTICLES = getStuckTrajectories(PARTICLES,
       minimum_kinenergy_threshold, N_stuck_steps)

%%IDENTIFIES AND FLAGS STUCK TRAJECTORY SEGMENTS (TO
   EXCLUDE THEM FROM DIFFUSIVITY
%CALCULATIONS)

% Brownian diffusion relies on the assumption of freely
   moving particles, which is
% no longer satisfied when a nanoparticle gets stuck/
   inmobilized in a nanochannel.
% Ideally, the kinetic energy is zero when a particle is
   not moving.
% However, thermal fluctuations still play a role,
   causing the position to
% vibrate around an equilibrium position. Therefore some
   sore of smoothing
% is necessary to detect a stuck trajectory.

%% ALGORITHM
%   * Smooth the trajectory with a moving mean
%   * Calculate a normalized (to the mean) kinetic energy
    per particle.
%   * Establish a minimum kinetic energy threshold to be
   considered as zero,
%   thus, a particle is stuck, not moving.
%   * Flag the segments of kinetic energy below the
   threshold and save them
%   into a new structure contained in PARTICLES.

%% INPUTS:
```

```
% PARTICLES data structure, more precisely, PARTICLES.
   timeFrame and
% PARTICLES.position

%% OUTPUTS:
%PARTICLES(i).StuckTrajectories for the i-th particle.
   Each of these is a data
%structure itself, containing timeFrame and position for
   each stuck segment.

% Call the n-th stuck segment of the i-th particle
% timeFrame = PARTICLES(i).StuckTrajectories(n).timeFrame
% position = PARTICLES(i).StuckPosition(n).position

%% PARAMETERS
%   - minimum_kinenergy_threshold = 0.05; % Threshold
   ratio of kinetic energy (with respect to the mean) to
   be considered as zero. To be refined according to
   stochastic processes theory
smoothing_parameter = 0.05; % Ratio of total number of
   frames of a trajectory to be taken into account for
   smoothing the position. Default is 0.05 = 5%
%   - N_stuck_steps = 5; Minimum number of time steps
   that a trajectory is allowed to
%       be in a stuck mode. The probability of this event
    is
%       [CDF(minimum_kinenergy_threshold)]^N_stuck_steps,
    where CDF is the
%       cumulative distribution function of the chi-
   squared distribution with k=1
%       degree of freedom


%% Calculate kinetic energy for each trajectory in
   PARTICLES struct
for particle = 1: length(PARTICLES)

    time = PARTICLES(particle).timeFrame; % Extract the
       time from the data frame

    % Number of frames to smooth up the trajectory.
       Rounded to the closest
    % odd integer.
    smoothing_frames = ceil(smoothing_parameter * length(
       time)); % Rounded up to closest integer
```

```
% Ensures an odd sliding averaging window for a
   centered-value mean
smoothing_frames = smoothing_frames + 1 - mod(
   smoothing_frames, 2);
% Sets a limit of frames in order to prevent over-
   smoothing
if smoothing_frames > N_stuck_steps
    smoothing_frames = N_stuck_steps; % Multiply by 5
        ms to get time in seconds
end

%Extract refined position if available
if isfield(PARTICLES, "positionRefined")
    position = PARTICLES(particle).positionRefined;
else
position = PARTICLES(particle).position; % Extract
   the position from the data frame
end

% Calculates kinetic energy divided by m/2, since
   dimensions are not important,
% but its relation to the mean, given by the chi-
   squared distribution

mean_kin_energy = mean( (diff(position).^2) );

% Denoise position with a moving mean
position = movmean(position, smoothing_frames); %
   Apply a moving mean with a sliding averaging
   window

% Kinetic energy normalized to the mean (
   dimensionless)
kin_energy = diff(position).^2;
kin_energy = kin_energy / mean_kin_energy;
kin_energy = movmean(kin_energy, smoothing_frames); %
    Apply a moving mean to the square of velocity

% Finds the times when the kinetic energy is below
   the minimum kinetic energy threshold. Flag=1 when
   kin energy is less than the threshold, 0 otherwise
stuck_times = kin_energy <
   minimum_kinenergy_threshold; % Logical array

%Find the starting and ending points of the stucking
   segments.
```

```
start_indices = find( diff([0, stuck_times.']) == 1);
    % Finds when a 1 is after a 0: Stucking starts
end_indices = find( diff([0, stuck_times.']) == -1);
   % Finds when a 0 is after a 1: Stucking ends

% Checks if trajectory was initiated in stuck mode.
   If so, sets the
% initial start index to the initial time. After
   these conditional
% statements, start_indices and end_indices must have
    the same length
if ( ~isempty(start_indices) ) && ( ~isempty(
   end_indices) ) % Check that at least one stuck
   segment has been found

    if end_indices(1) < start_indices(1)
        start_indices = [time(1), start_indices];
    elseif end_indices(end) < start_indices(end) % If
       trajectory ends in stuck mode, sets the final
       end index to the final time
        end_indices = [end_indices , length(time)];
    end

    % FOR DEBUGGING: Flag if lengths do not coincide.
    if length(start_indices) ~= length(end_indices)
        disp("getStuckTrajectories SUBROUTINE ERROR:
           ")
        disp("START INDICES' LENGTH DO NOT COINCIDE
           WITH END INDICES")
    end


% ASSIGN TIME AND POSITION OF THE STUCK SEGMENTS INTO
    A NEW STRUCTURE
    stuck_segments = length(start_indices); % Number
       of times a given particle/trajectory gets
       stuck
    % Counter or segments which are longer than the
       smoothing_frames
    % parameter
    count = 0;
    for n = 1 : stuck_segments
        %disp(n)
        % Avoid counting turning points smoothed out
        if end_indices(n) - start_indices(n) >=
           N_stuck_steps
```

```
            % Fetch time and position for each stuck
               segment of each trajectory
                particle

                % Corrects with an extra time step, for
                   the discrete trajectory being defined
                   as a forward finite difference.
                if end_indices(n)<length(PARTICLES(
                   particle).timeFrame) %Ensures index
                   contained within trajectory
                    end_indices(n) = end_indices(n) + 1;
                elseif start_indices(n) > 1
                     start_indices(n) = start_indices(n)
                        - 1;
                end

                time = PARTICLES(particle).timeFrame(
                   start_indices(n) : end_indices(n) );
                position = PARTICLES(particle).position(
                   start_indices(n) : end_indices(n) );

                % Number of times a particle gets stuck
                count = count + 1
                PARTICLES(particle).StuckTrajectories(
                   count).timeFrame = time;
                PARTICLES(particle).StuckTrajectories(
                   count).position = position;
                %disp( strcat("Particle analyzed: ",
                   num2str(particle), ", number of stuck
                   segments:  ", num2str(count)) )
            end
        end
    end
end
end
=
```

DEPARTMENT OF PHYSICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY