LUNDS UNIVERSITET

Derivation of an Algorithm for the Analysis of Images of DNA molecules prepared with Denaturation Mapping

MAXIMILIAN BINZLER

Thesis submitted for the degree of Master of Science
Project Duration: 10 months

Supervised by

PROF. JONAS TEGENFELDT, DR. JASON BEECH and OSKAR STRÖM

Department of Physics
Division of Solid State Physics
May 2019

Abstract

Denaturation mapping is a powerful and fast method for optically analyzing DNA molecules. It can be used to characterize DNA molecules on a scale of a few hundred base pairs which is sufficient for applications. Denaturation mapping requires a sufficiently high optical resolution of the imaging system used, which typically translates to an expensive, bulky microscope. For point-of-care testing or usage in poor, remote regions of the world, it would be desirable to perform denaturation mapping without the need of a high-end microscope. In this thesis, an algorithm is presented and analyzed which reconstructs a high-resolution barcode signal from a series of low-resolution images. The implementation of this algorithm in imaging systems with optical resolutions that are too low to be useful otherwise could unlock the possibility of performing denaturation mapping. A necessary precondition for the algorithm to work is that the series of low resolution images have a known, unidirectional shift between them. With this knowledge, the signals of the different images can be merged into a single signal which can then be deconvolved with a square function of the same size as a single pixel. This would, in theory, lead to a resolution that is only dependant on the step size between the images, which is very suitable for imaging systems with large pixel sizes. Those imaging systems generally have the advantage of having a better signal to noise ratio than imaging systems with smaller pixel sizes. The aim was to provide a proof-of-principle of the idea behind the algorithm and to verify it with experiments. During this master thesis the algorithm was tested on artificial data and afterwards, its performance was analyzed with over 60 DNA molecules.

List of Abbreviations

DNA Deoxyribonucleic acid

A-T Adenine-Thymine (bond)

C-G Cytosine-Guanine (bond)

fps Frames per Second

OFM Optofluidic Microscopy

SROFM Sub-pixel resolving Optofluidic Microscopy

YOYO-1 Tetracationic homodimer of Oxazole Yellow

QPM Quantitative Phase Microscopy

LED Light Emitting Diode

CMOS Complementary Metal-Oxide-Semiconductor

CCD Charge-Coupled Device

EMCCD Electron Multiplying Charge-Coupled Device

ITU International Telecommunication Union

BME Beta Mercapto Ethanol

TBE Tris Borate EDTA buffer

EDTA Ethylenediaminetetraacetic acid

MES 2-(N-morpholino)ethanesulfonic acid

NA Numerical Aperture

bp base pair

PSF Point-Spread-Function

FOV Field Of View

cmr cooperatively melting regions

lcm lowest common multiple

mod modulo operator

RMSE Root Mean Square Error

std standard deviation

SNR Signal-to-Noise

PCR Polymerase Chain Reaction

MAF Moving Average Filter

List of Variables

l_{DNA} Length of the DNA Molecule [m]

f_{sample} Sampling Frequency [1/s]

l_{pixel} Length of a Pixel [m]

v_{DNA} Speed of the DNA [m/s]

l_{total} Total Distance covered by the DNA during the measurement [m]

n_{step} Number of Frames of the Video Sequence [-]

n_{pixel} Number of pixels to track the DNA movement [-]

l_{step} Distance covered by the DNA between two adjacent Frames [m]

l_{sample} Sampling Distance in the evenly sampled region of the Signal [m]

$t_{shutter}$ Off-time of the shutter between two adjacent frames [s]

f_{cmr} Inverse of the length of the cooperatively melting regions of the DNA molecule [1/m]

l_{step}^{lb} Lower bound on the step size [m]

$n_{step,max}$ Number of steps before the DNA is imagined in the same way as before [-]

$n_{pixel,min}$ Number of pixels the DNA covers during $n_{step,max}$ frames [-]

l_{start} Start of the evenly sampled region [m]

l_{end} End of the evenly sampled region [m]

$y_{linear,n}$ Linear Signal of the *nth* frame [-]

pos_n Position Vector of the *nth* frame [m]

y_{1D} One-dimensional Signal obtained after the pre-processing [-]

pos_{1D} One-dimensional Position vector obtained after the pre-processing [m]

$f(x)$ Analog DNA barcode [m]

H Low resolution Image

W High resolution Image

S Point Spread Function

$g(r)$ Gaussian Approximation to the Airy function

$f_{blur}(x)$ Reconstructed Signal before the deblurring step

$f_{recon}(x)$ Reconstructed Signal after the deblurring step

f_{box} Size of the low-pass fourier filter [1/m]

Contents

Chapter 1

Introduction

Since its invention, the basic setup of a conventional microscope has not changed significantly. Although the capabilities of microscopes and the resolutions achieved have increased tremendously, the basic setup has stayed more or less the same. This setup with its bulky, fragile and expensive parts, like the objective or the eyepiece, can pose big problems. This is especially so for point-of-care testing and use in remote, poor regions of the world. It is therefore desirable to have a microscope setup that eliminates all (or most) of the bulky, expensive parts while maintaining the high resolution necessary for biological or medical imaging [1].

There are several approaches to slimming the bulky, expensive microscope design while maintaining high resolution. One of them is called optofluidic microscopy (OFM) or sometimes sub-pixel resolving optofluidic microscopy (SROFM). OFM uses a tilted array of apertures that is put directly onto a sensor grid. Figure 1.1 shows the mechanism of the OFM method by showing the different steps, from a simple, direct-projection image, to a filled-in, raster-scan image that is achieved by translation of the sample. By moving the sample over this tilted array of apertures, a filled-in, high-resolution image is created from a series of low resolution, direct-projection images. This method has, for example, been used by Cui et al. [2] and Heng et al. [3] to image *Caenorhabditis elegans*, for which they achieved a resolution of 490 ± 40 nm. The apertures used by this method are smaller than the wavelength of the light used, which makes the physics of light transmission not straightforward. That is why Heng et al. analyzed the transmission in [4].

Bishara et al. [5] used partially coherent, in-line holography (HOM) as an imaging approach for optofluidic microscopy. With this they achieved a resolution of $\sim 1\,\mu$m. Zheng et al. [6] and Lee et al. [7] presented a method where they created a high resolution image with a res-

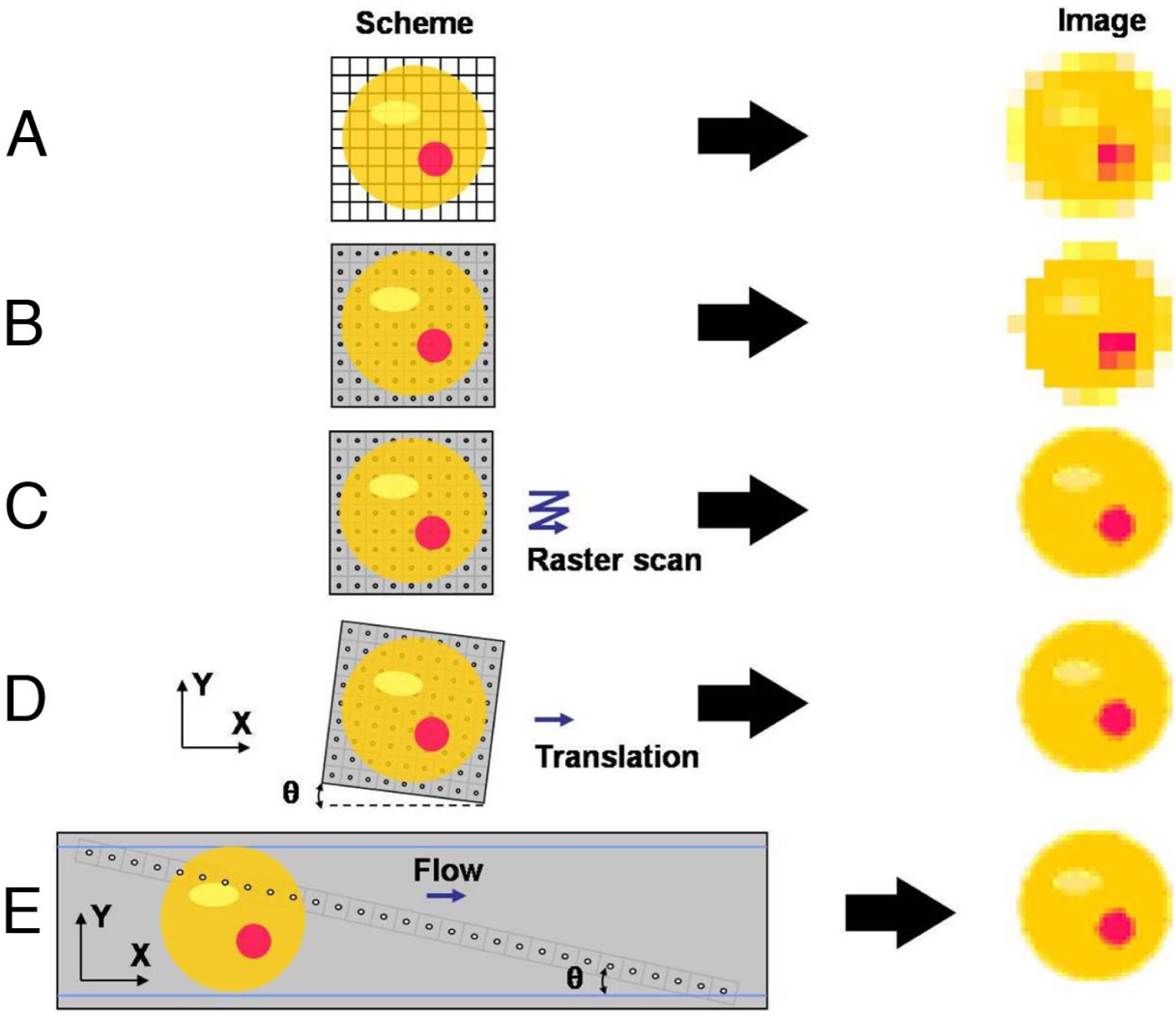

Figure 1.1: Mechanism of the OFM method, figure adapted with permission from [2]. A: Direct Projection Image B: Sparsely sampled images created by putting apertures on the sensor-grid C: Raster-scanning creates a filled-in image D: Translation of the sample can replace the raster-scan E: The aperture grid can be replaced by a long, linear array.

olution of ∼660 nm from a series of undersampled, low resolution images. Done with a setup that abandons the tilted aperture array used in the aforementioned OFM works, they achieved an increase in resolution beyond the pixel size limit by using an algorithm which goes back to works of Farsiu et al. [8].

Another approach that abandons the bulky setup of a conventional microscope is quantitative phase microscopy (QPI). For QPI, the optical path-length shifts of the light that is transmitted through the sample are measured which contain information about the thickness and the refractive index of the sample. The path-length shifts are detected with interferometry. Guo et al. [9] gives a good overview of the method and its various applications such as holography [10],

dispersion-relation phase spectroscopy [11], Fourier-transform light scattering [12] and time-stretch QPI [13][14]. Time-stretch QPI is an especially interesting method for point-of-care testing and use in poor, remote regions since it is easy to combine with microfluidics. Lau et al. [14] reported a resolution of ∼1.2 µm for their method and Guo et al. [13] reported a resolution of 980 nm.

Another approach towards abandoning the bulky setup of a conventional microscope is to

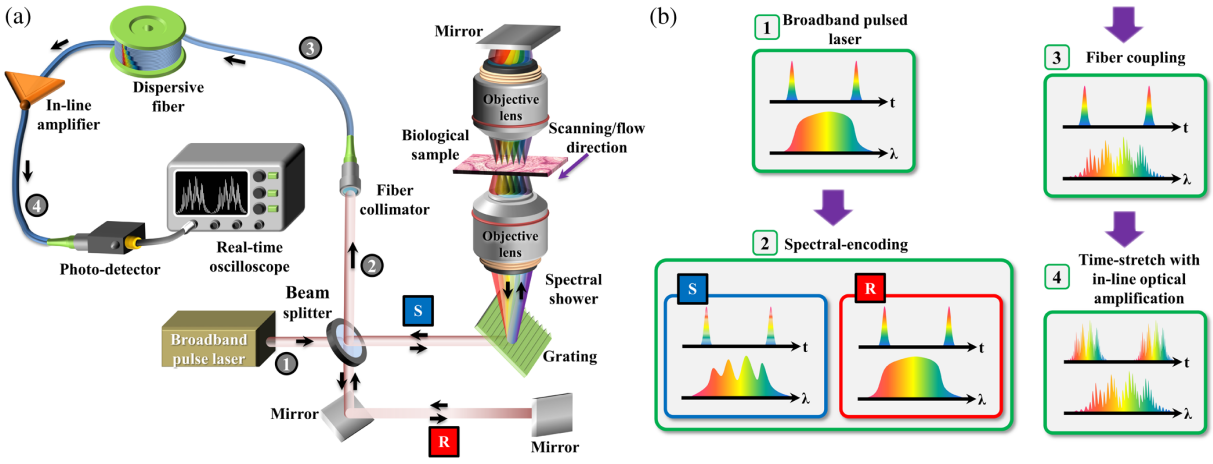

Figure 1.2: Mechanism of the QPI method. Figure adapted with permission from [14]

use a smartphone as a microscope. Koydemir and Ozcan said that in 2017 the International Telecommunication Union (ITU) reported ∼7.7 billion smartphone subscriptions, with 10% of these subscriptions being located in the least-developed countries [15]. Considering such an abundance, using a smartphone as a medical imaging device could improve global health-care tremendously. There are several different approaches to using a smartphone as a microscope. Tseng et al. [16], for example, used an attachment with a simple LED to create a lensfree hologram of the sample being imaged. They reported a resolution of 2.2 µm which was limited by the pixel size of the CMOS sensor used in the camera. Lee et al. [17] reported a method which abandons all additional devices and built-in light sources. They removed the lens module of the camera and put the sample directly onto the sensor grid. For the method to work, any external light source such as the sun or a conventional lamp could be used. They reported that they were able to resolve 500 nm spheres, although they observed some aliasing. Kheireddine et al. [18] used a second smartphone as an illumination light source with different illumination modes. They reached an optical resolution of less than 2 µm. Other smartphone tools enable fluorescent imaging. For example, Wei et al. [19] presented a tool with which they were able to detect 100 nm sized beads, while Koydemir et al. [20] presented a smartphone tool that was used for

Giardia lamblia detection in water. A desirable goal for these microscopy methods is to enable DNA imaging, which has, for example, been done by Wei et al. [21] with a smartphone fluorescence tool on DNA origami structures.

A powerful method for the optical mapping of DNA molecules is called *denaturation mapping*. This method can be used, for example, for the detection of structural variations in DNA molecules or for the identification of pathogens like viruses and bacteria. The details of how this method works are described in Section 2.1. It provides kilobasepair resolution and was used by Nyberg et al. [22] to analyze DNA of the phage T4 and λ-DNA. Reisner et al. [23] also used denaturation mapping to analyze λ-DNA, T7, T4GT7 and a BAC construct from chromosome 12.

The work presented in this master's thesis should be seen in the context of the different projects described above. In this work, several series of images of fluorescently labeled, barcoded DNA molecules (see Section 2.1) were taken with a controlled, sub-pixel shift between every two adjacent frames of the image stack. By readjusting the images and deconvolving them with a square function of a size corresponding to the pixel size a high resolution barcode signal was obtained. This was done for different pixel sizes to evaluate the capability of the method. The goal of this work was to develop the algorithm and show its general functionality. This was a first step towards implementing the algorithm in imaging systems, such as smartphone microscopes, with pixel sizes too large to resolve the features displayed by the DNA barcode pattern.

This work is structured as follows: In chapter 2 the sample-DNA used in this work is presented as well as the method of creating a barcoded DNA molecule by denaturation mapping and the equipment used for the imaging process. In the following chapter the theoretical background of the algorithm used to obtain a high resolution image from a series of low resolution images and its mathematical foundation are explained. In chapter 4 the results of both the theoretical simulation and the experimental work are presented. Chapter 5 concludes this work and discusses difficulties and possibilities for future work.

Chapter 2

Experimental

2.1 Denaturation mapping

In this section, the two steps necessary to create a barcoded DNA molecule are explained. The first step is to prepare the DNA molecules with a technique called denaturation mapping. The second step is to elongate the prepared DNA molecules which can, for example, be achieved by stretching it out on a surface or by flowing it through a sufficiently small nanochannel. Denaturation mapping makes use of the different temperatures at which the different kinds of bonds in a DNA molecule, A-T and C-G, break. This process of bonds breaking is called melting of the molecule. As described by Yakovchuk et al., [24] the stability of a DNA molecule is mostly determined by base-stacking interactions. They show that at least half of the stability difference between A-T and C-G is due to heterogeneity of stacking interactions, whereas the other half is due to a difference in the energetics of A-T and C-G base pairing. These factors cause regions with a high concentration of A-T to melt before regions with a high concentration of C-G. The melting temperature of a DNA molecule can be lowered by adding formamide to the solution of DNA molecules. The formamide drops the melting temperature by lowering the ionic strength in proximity to the DNA molecule and by binding to broken bonds. This prevents them from closing again. The melting temperature for *E.Coli* genome without formamide is 83.5 °C and the melting temperature (in Celsius) drops in dependence of the formamide content according to $\Delta T_m = -0.6 \cdot F$, where F stands for the formamide concentration in percent [25]. The main DNA labeling techniques that can be used in combination with optical DNA mapping can be divided into two categories: *Enzyme Based Labeling* and *Affinity Based Labeling*. Denaturation mapping falls into the category of affinity based labeling. Müller and Westerlund give a good

overview of optical DNA mapping techniques in [26].

The DNA molecules were analyzed with a fluorescence microscope (see Section 2.3). Therefore, in order to get a fluorescent signal from the DNA molecules, it was necessary to add a fluorescent dye to them. The dye used in this master thesis was *YOYO-1*, which absorbs light at a peak-wavelength of 489 nm and emits light at a peak-wavelength of 509 nm when attached to the DNA molecule, and at 458 nm and 564 nm respectively when detached from the DNA molecule [27]. Figure 2.1 shows the excitation and emission spectrum of *YOYO-1* when the dye molecule is attached to the DNA molecule. The data for the spectra have been obtained from *Thermo Fisher Fluorescence SpectraViewer* [28]. The dye *YOYO-1* is basically non-fluorescent

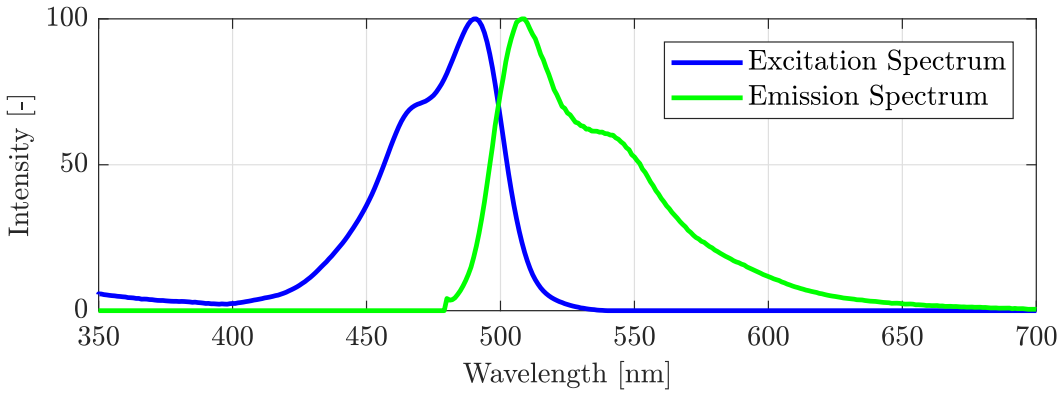

Figure 2.1: Excitation and Emission spectra of *YOYO-1* when attached to the DNA molecule [28].

when it is not attached to the DNA molecule [27]. When the DNA molecule is partially denatured in the way described above, the dye molecules will detach from the broken bonds, thereby letting regions with a high concentration of A-T bonds appear dark in a fluorescent image and areas with a high concentration of C-G bonds appear bright.

To make this so-created barcode pattern visible, it is necessary to linearly elongate the DNA molecules. This can be done by flowing them through a nanochannel or by stretching them out on a surface. The surface needs to be either positively charged, to attract the negatively charged backbone of the DNA molecule, or hydrophobic [29]. In this work, the DNA molecule was stretched out by dragging a droplet over a surface covered with a polymer called *Zeonex*®. This method has been described by Deen et al. [29]. The base pair resolution of this method is limited by the number of cooperatively melting regions (cmr) along the DNA sequence, which is several 100 bp [23]. This resolution is further degraded by the imaging system (optical diffraction, finite pixel size).

In this work, the DNA barcodes were created by first mixing 10 µL of the DNA stock (10 µm λ-DNA in 5x TBE stained with *YOYO-1* at a ratio of 10:1 stained bp:YOYO-1) with 10 µL of formamide. The mixture was heated at 45 °C in a PCR (*Polymerase Chain Reaction*) machine for 10 minutes and afterwards transferred into 80 µL of an ice-cooled buffer solution (50 µL of 100 mM MES-buffer mixed with 30 µL Milli-Q water). To stretch the DNA molecules out on a surface, a droplet of the DNA solution was put onto a *Zeonex*® cover slide and brought into contact with a pipette-tip, (that was wrapped with *Parafilm* in order to avoid the droplet being sucked into the pipette-tip). Next, it was dragged over the surface by the pipette-tip, causing the DNA molecules to get stretched out on the surface.
For the analysis of the DNA molecules, a fluorescent microscope (Nikon TE2000-U) was used. It had a 512×512 EMCCD camera (iXon 897 from *Andor Technology*) with a pixel size of 16 µm. The objective used for the imaging was a 100x magnification Plan Apo objective with a NA of 1.4. This was an oil-immersion objective which required an immersion oil with a refractive index of $n_D = 1.515$. The light source used by the microscope was a *lumencor* SOLA 6-LOR-SB white light source.

2.2 Analyzed Sample

The DNA molecules analyzed in this master thesis were from the bacteriophage *Enterobacteria phage* λ, which infects *Escheria coli* (E.coli) bacteria. In its circular form, the DNA molecule is 48502 base pairs long which is roughly equal to a contour length of 16.5 µm, assuming a base pair length of 0.34 nm. Upon loading with the *YOYO-1* dye used in this work, the length of the molecule increases up to 22 µm, 1 dye molecule per 4 bp [30], thereby increasing the contour length of the molecule by 0.68 nm per dye molecule [31]. In this master thesis, a ratio of 1 dye molecule per 10 base pairs was used. It was empirically found that this ratio leads to the best results. Furthermore, the dye molecules unwind the helical structure of the DNA molecule [31]. Lambda-phage DNA has an equal amount of C-G and A-T bonds [32]. Figure 2.2 shows the barcode pattern of lambda-phage DNA.

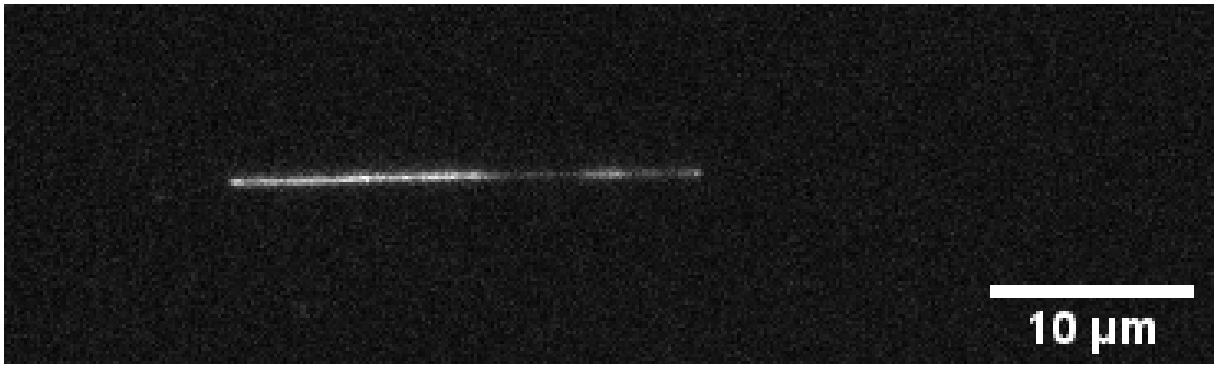

Figure 2.2: Barcoded Lambda-phage DNA

2.3 Fluorescent Imaging

As mentioned in Section 2.1, the microscope used in this master thesis was a Nikon TE2000-U. The microscope was used as an epi-fluorescence microscope meaning that the excitation and the detection happened both on the same side of the sample as opposed to dia-illumination where the excitation and detection happen on opposite sides of the sample. Figure 2.3 shows schematically the functionality of an epi-fluorescence microscope with its most important parts. The excitation light travels from the light source through the excitation filter to the dichroic mirror where it is reflected and afterwards focused onto the sample with the objective. The sample then absorbs the excitation light and re-emits light of a longer wavelength. The emitted light then travels back through the objective to the dichroic mirror where it is, in contrast to the excitation light, not reflected but transmitted. After passing through the dichroic mirror the emitted light passes through the emission filter and into the detector. The excitation and emission filters are necessary to make sure that only the excitation light reaches the sample and that only the emitted light reaches the detector. This is necessary because the dichroic mirror is not very efficient, it transmits around 20 % of the excitation light.

As mentioned in Section 2.1, *YOYO-1* absorbs light at a peak-wavelength of 489 nm and emits light with a peak-wavelength of 509 nm when attached to the DNA molecule; therefore a filter-cube was adjusted to these wavelengths. In this thesis, a *FITC* filter cube was used. Its excitation filter had a mean pass-wavelength of 482 nm with a bandwidth of 35 nm and its emission filter had a mean pass-wavelength of 536 nm with a bandwidth of 40 nm. The dichroic mirror used in this filter cube reflected light of wavelengths below 505 nm and transmitted wavelengths above this value.

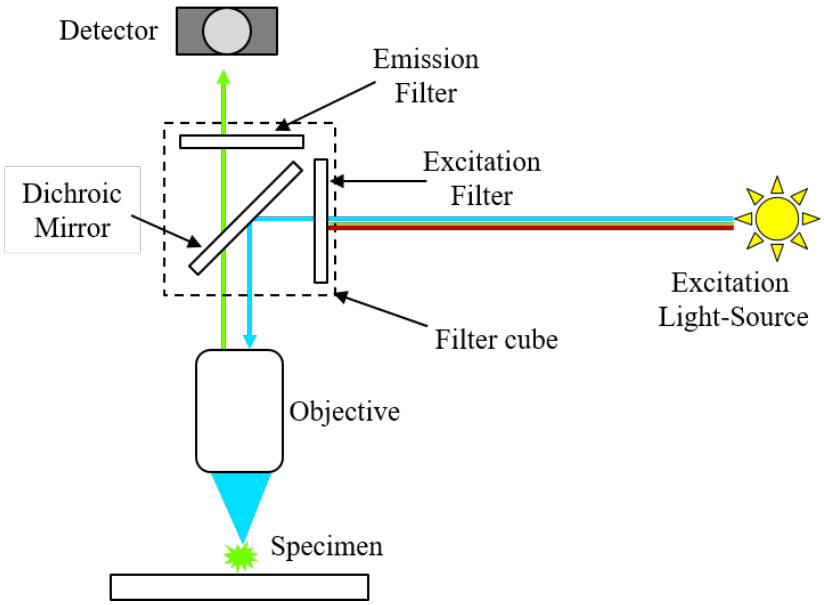

Figure 2.3: Schematic of an epi-fluorescence microscope.

2.4 Imaging sensor

The images captured in this work were all obtained by using a camera with an EMCCD sensor grid where EM stands for *electron multiplying* and CCD stands for *Charge Coupled Device*. This section will first explain the functionality of a CCD sensor, then briefly address the electron multiplying effect.

2.4.1 Functionality of a CCD sensor

The architecture of a CCD sensor is shown in figure 2.4. The CCD sensor in figure 2.4 is a back-illuminated CCD sensor like the one used in this master thesis. Back-illumination means that, in contrast to a front-illuminated sensor, the light does not have to pass through the wiring or the gate electrodes. Back-illumination provides a higher sensitivity, especially for shorter wavelengths [33]. A CCD sensor converts the energy of an incoming photon into an electron-hole pair in the silicon underneath the gate. The CCD then collects the produced electrons in so called ”potential wells” which correspond to the pixels, whereas the holes are displaced in the silicon substrate. The number of electrons in a potential well is proportional to the amount of photons that have reached a particular pixel. To convert these photogenerated electrons into a digital pixel value, the CCD gate voltages are modified to transfer the electrons to a ”read-out” circuit where they are converted to a digital signal [33]. The CCD camera used in this master thesis was *ANDOR iXon DV887* with a field of view of $8.2\times$ $8.2\,\text{mm}$ with 512×512 pixels

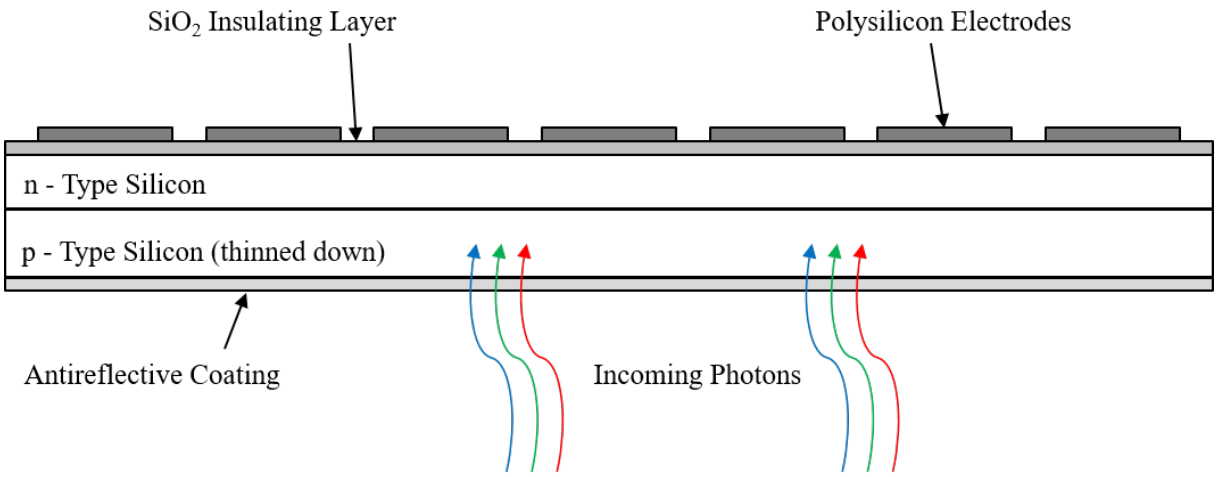

Figure 2.4: Cross-section of a back-illuminated photogate-based area-array CCD (modified from Fig. 3.7. page 36 in [33])

leading to a pixel-size of $16 \times 16\,\mu m$. To reduce thermal noise the camera was cooled down to a temperature below $-80\,^{\circ}C$.

2.4.2 Functionality of the electron multiplying effect

The difference between a conventional CCD sensor and an EMCCD sensor is that a series of multiplying registers are added in front of the read-out circuit of the conventional CCD setup which amplifies the electron signal. The effect these multiplying registers use is known as *impact ionization* [34]. During impact ionization, a charge-carrier with sufficient kinetic energy can cause the generation of other electron-hole pairs. If every charge-carrier produced in this manner creates more than one electron-hole pair, the current at the end of the multiplying registers is amplified. The necessary kinetic energy is provided by a strong electric field in the multiplying registers. Another name for this process is *avalanche multiplication* [35]. The advantage of this amplification is that it essentially eliminates read-out noise, allowing much higher read-out rates and detection of very small signals [36].

Chapter 3

Theory

In this section, the theory behind the algorithm used to create a high resolution DNA barcode signal from a series of low resolution images is explained. In the first section of this chapter, the variables necessary to describe the algorithm are introduced. The second section will explain the mathematical basis for the algorithm, neglecting detrimental effects such as noise and blur. Section three will address these problems and their influence on the performance of the algorithm.

3.1 Definitions of the Variables used in the Algorithm

In order to get a one-dimensional signal from the stack of frames, it was necessary to apply a few pre-processing steps. The first one was to choose an appropriate FOV. Figure 3.1 shows an appropriate FOV for a few different frames. The FOV should cover the whole movement of the DNA molecule which means that the FOV needs to be large enough to show the whole DNA molecule all the way from the first frame to the last frame. On the other hand the FOV should be chosen to be as small as possible to avoid the contribution of noisy pixels to the signal. The more the DNA molecule is aligned with the direction of movement the smaller the FOV can be chosen and the better the algorithm works.
Figure 3.1 shows schematically the pre-processing steps. On the left side of figure 3.1, the first, second and last frame of the image stack are shown, On the right side, the summation process and the resulting barcode after the algorithm has been applied is shown. On the basis of this, the variables necessary to describe the imaging process are defined: The length of the DNA molecule is denoted as l_{DNA} while the length of one pixel is denoted as l_{pixel}. The number of

pixels necessary to cover the entire movement of the DNA molecule is denoted as n_{pixel} and the total number of frames in the imaging process is denoted as n_{step}. The distance the DNA molecule covers between two adjacent frames is denoted as l_{step}. If the stack of low-resolution images necessary to reconstruct the high-resolution image is taken by recording every image in the stack separately and moving the DNA molecule between adjacent frames with a platform, then this quantity is given directly. If the DNA molecule is flowing in a nanochannel with a certain velocity v_{DNA} however, then the length covered between two adjacent frames can be calculated by multiplying this velocity with the off-time of the shutter of the camera $t_{shutter}$. The total distance the DNA molecule covers during the imaging process is denoted by l_{total}, which can be calculated according to $l_{total} = (n_{step} - 1) \cdot l_{step}$.

After selecting the FOV, the pixel values in the FOV are summed along the direction perpen-

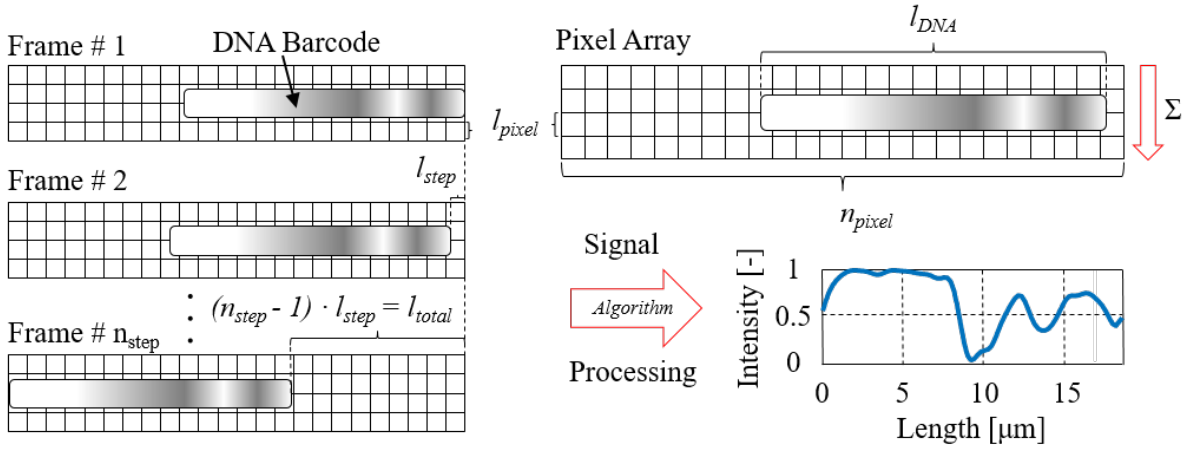

Figure 3.1: Variables necessary to describe the image processing: *Top Left:* First Frame, *Middle Left:* Second Frame, *Bottom Left:* n_{step}-frame; *Top Right:* Summing process that was applied for every frame, *Bottom Right:* Resulting barcode Signal

dicular to the direction of the movement of the DNA molecule. This is done for every frame and is schematically shown in the top right of figure 3.1. This creates a vector for every frame with a length equal to the number of pixels n_{pixel} in the direction of movement of the DNA molecule. The vector for the nth frame is denoted by $y_{linear,n}$. To correctly process the signal, it was necessary to assign this intensity value to a corresponding spatial value. This was chosen such that the first value in the vector of the first frame, $y_{linear,1}$, had the length coordinate zero; the second value of $y_{linear,1}$ had the length coordinate l_{pixel}; the third value of $y_{linear,1}$ had the length coordinate $2 \cdot l_{pixel}$ etc., for the first frame. After that, the first value in the vector for the second frame, $y_{linear,2}$, was assigned the coordinate l_{step}; the second value in $y_{linear,2}$ was

assigned the coordinate $l_{pixel} + l_{step}$ etc. In general, the mth value in the vector of the nth frame was assigned the length coordinate $(m-1) \cdot l_{pixel} + (n-1) \cdot l_{step}$. Similarly, a position vector for every frame was created where the vector for the *nth* frame was denoted as pos_n.
These so-created n_{step} intensity- and length-vectors were then each put into a single 1D-vector with a length of $n_{pixel} \cdot n_{step}$, which were denoted as y_{1D} and pos_{1D} respectively. Afterwards, pos_{1D} was sorted so that the spatial values were monotonically increasing. After that, the vector y_{1D} was sorted accordingly. This created signal was the basis for the algorithm presented in Section 3.2.

3.1.1 Artificial Signal

To evaluate the algorithm, it was helpful to have a noise-free, artificial, sample data set to test it on. This artificial barcode was created by an algorithm from Tobias Ambjörnssons group from Lund University and is based on the known DNA sequence of λ-phage. The barcode and the one-sided power spectrum of its Fourier transform are shown in figure 3.2 (without the DC-component). Throughout the next section this barcode is used to evaluate the algorithm. The intensity profile shown in figure 3.2 will is denoted by $f(x)$, and its Fourier transform, $\mathscr{F}(k)$, where k denotes spatial frequencies which have the SI unit of $[\frac{1}{m}]$.

3.2 Derivation of the Algorithm

Two assumptions that are made throughout this section are, firstly, that there is no noise corrupting the signal. and secondly, that there is no blur from the imaging system corrupting the signal ($\rightarrow \delta$-function as PSF). As mentioned in Section 2.1, the method of denaturation mapping has an inherent resolution limit which is due to the cmr's in the DNA molecule. This resolution limit is on the order of several 100 base pairs, about 50 nm. Frequencies much higher than the inverse of this length can therefore be regarded as noise. This is congruent with the power spectrum shown in figure 3.2 which drops rapidly for higher frequencies. Any algorithm that aims at increasing the resolution of a barcoded DNA signal should not try to restore features that are below this limit. If the frequency corresponding to this inherent resolution limit is denoted by

$$f_{cmr} = \frac{1}{l_{basepair} \cdot cmr} \tag{3.1}$$

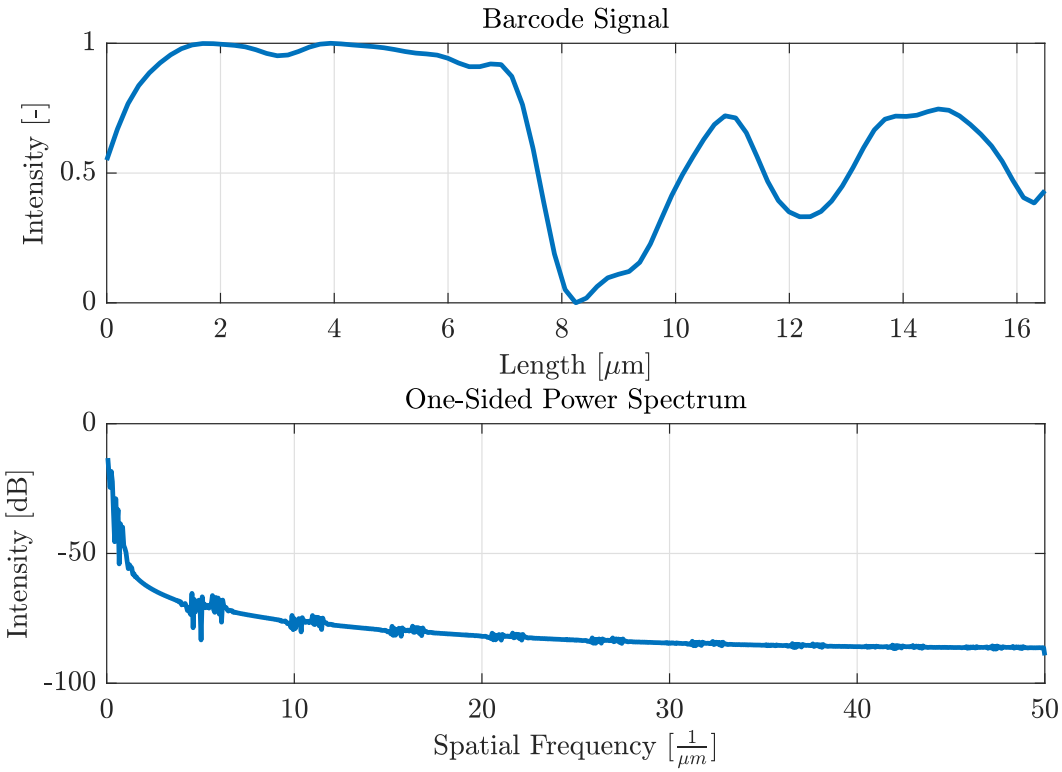

Figure 3.2: Artificial barcode used to derive the algorithm and its Fourier transform.

then, according to the Nyquist-Shannon sampling theorem [37][38], the sampling frequency does not need to be higher than

$$f_{sample} = 2 \cdot f_{cmr} \tag{3.2}$$

One must be careful here. The Nyquist-Shannon theorem does not directly apply because a pixel does not take the value of a single point of the analog DNA barcode $f(x)$ but rather the sum over a whole pixel length. This summing process is equivalent to convolving the signal with a square function. This means that the measured signal is a convolution of $f(x)$ sampled at a certain sampling frequency, and a square function with a size equal to one pixel. It is therefore necessary to first deconvolve the signal. If a perfect deconvolution is assumed, then the result of the deconvolution is exactly the analog DNA barcode $f(x)$ sampled with the aforementioned sampling frequency. This will be addressed in more detail below.

The reciprocal of this sampling frequency gives a good estimate of a lower bound on the step size between two adjacent frames:

$$l_{step}^{lb} = \frac{1}{f_{sample}} = \frac{l_{basepair} \cdot cmr}{2} \tag{3.3}$$

The maximum number of steps for a given step-size and a given pixel size can be calculated according to:

$$n_{step,max} = \frac{lcm(l_{pixel}, l_{step})}{l_{step}} \tag{3.4}$$

where *lcm* stands for the *lowest common multiple*. The *lowest common multiple* is the smallest number that is divisible by both of the input numbers. For example, the *lcm* of 4 and 6 is 12.

After $n_{step,max}$ number of steps the barcoded DNA molecule has moved

$$n_{pixel,min} = \frac{lcm(l_{pixel}, l_{step})}{l_{pixel}} \tag{3.5}$$

pixels and the process repeats itself, measuring the barcoded DNA molecule in the same way as $n_{step,max}$ steps before. It is therefore, theoretically, unnecessary to take more frames n_{step} than $n_{step,max}$. In the region starting from the point

$$l_{start} = (n_{pixel,min} - 1) \cdot l_{pixel} - l_{step} + gcd(l_{pixel}, l_{step}) \tag{3.6}$$

$$= (n_{pixel,min} - 1) \cdot l_{pixel} - l_{step} + l_{sample} \tag{3.7}$$

to the point

$$l_{end} = (n_{pixel} - 1) \cdot l_{pixel} + n_{step,max} \cdot l_{step} - l_{start} \tag{3.8}$$

the barcoded DNA signal is sampled with a step-size of $l_{sample} = gcd(l_{pixel}, l_{step})$, where *gcd* stands for the *greatest common divisor*, which is the greatest number that divides both l_{pixel} and l_{step} in this case. For example, the *gcd* of 12 and 9 is 3.

It is now worth thinking about how long the distance from l_{start} to l_{end} should be. Every point along this distance is the sum over a window of the size of l_{pixel} of the DNA molecule which corresponds to a convolution of the DNA signal with a square function equal to the pixel size. To correctly reconstruct the actual DNA signal, the distance from l_{start} to l_{end} should cover the *full* convolution of the actual DNA signal and the square function. That means:

$$l_{end} - l_{start} > l_{DNA} + l_{pixel} \tag{3.9}$$

Since l_{sample} is by definition smaller than l_{step} the following equation holds:

$$l_{start} < (n_{pixel,min} - 1) \cdot l_{pixel} \tag{3.10}$$

From equations 3.8, 3.9 and 3.10, the number of pixels ($\rightarrow$ FOV) necessary to track the movement of the DNA molecule can be estimated via:

$$n_{pixel} > 2 \cdot l_{start} + l_{DNA} + l_{pixel} \approx l_{DNA} + 2 \cdot n_{pixel,min} \cdot l_{pixel} \tag{3.11}$$

which is a more complicated way of saying that the FOV should extend sufficiently to both sides of the DNA molecule. This it does, trivially, since it needs to cover the movement of the DNA molecule over the whole movement.

A good thought experiment to see the equivalence of the imaging process described above and the barcode signal $f(x)$ shown in figure 3.2 getting convolved with a square function is to imagine instead of the DNA molecule moving over several pixel the DNA molecule to be stationary and a single pixel with a length of l_{pixel} moving with a step size of l_{sample} over the whole barcoded DNA molecule which is exactly a convolution of the barcoded DNA signal with a square function the size of the pixel.

If noise and blur are neglected, the best estimate of the barcoded DNA signal $f(x)$ sampled with a distance of l_{sample} is the deconvolution of the 1D-signal y_{1D} described in Section 3.1. The limit of how well this reconstructed signal will match the analog barcoded DNA signal depends only on the performance of the deconvolution algorithm evaluated in Section 4.1.

3.2.1 Deconvolution Algorithm

Since deconvolution is an ill-posed inverse problem, its solution is not trivial and there exists an extensive literature around solving deconvolution problems. MATLAB has a few built-in deconvolution algorithms, for example Wiener, Blind and Lucy-Richardson. In this master's thesis, the Lucy-Richardson deconvolution algorithm *deconvlucy* was used to reconstruct the barcode signal. The algorithm goes back to work done by William H. Richardson [39] and Leon B. Lucy [40]. The algorithm uses Bayes theorem iteratively to estimate the original signal from a degraded signal and a known PSF (in this case a square function). The algorithm was developed for two dimensional images but can be applied to one-dimensional signals without any problems. The implementation of the algorithm in MATLAB uses an accelerating, vector-extrapolation technique that was developed by David S.C. Biggs and Mark Andrews in [41].

A problem that has been encountered throughout this thesis was that the *Lucy-Richardson* algorithm lead to aliasing effects such as ringing at the beginning of the signal. The *Lucy-Richardson* algorithm assumes periodicity of the data [42], therefore the data was extrapolated with a spline function before the deconvolution step to ensure continuity and differentiability. Figure 3.3 shows two deconvolutions of the same signals: on the left without spline interpolation before the deconvolution, and on the right with spline interpolation before the deconvolution.

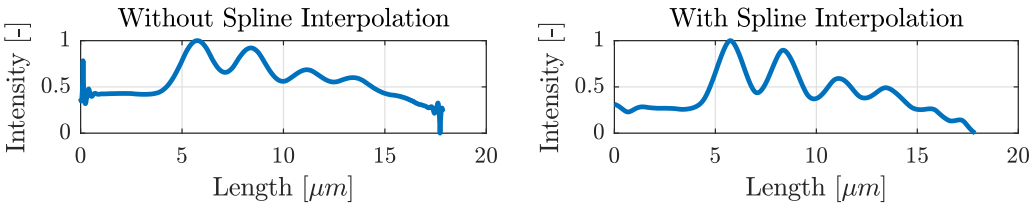

Figure 3.3: *Left*: Signal deconlved without interpolating with a spline function first. *Right*: Signal deconlved with interpolating with a spline function first.

The comparison of the two signals shows that the spline interpolation suppresses the ringing artifact which can be seen for the signal on the left but not for the signal on the right.

3.3 The effect of Noise and Blur

In contrast to the mathematical derivation, an experimental result will always contain both noise from different sources (e.g. Gaussian noise or thermal noise) and blur from the PSF of the imaging system. In this section, possible solutions for dealing with these detrimental effects are presented. First, blur due to the PSF of the imaging system is addressed, afterwards, noise.

3.3.1 Blur due to the PSF of the Imaging System

In every imaging system a single point in the object plane will be smeared out to a finite sized area in the image plane. This is due to the finite size of the aperture in the focal plane. Mathematically this process can be described by a convolution of the original image W with the PSF S which leads to the degraded image H which is observed:

$$H = W \circledast S \tag{3.12}$$

If the PSF is known, it is possible to increase the resolution of the blurred image H by *deconvolving* it with S. As mentioned in Section 3.2.1, deconvolution is an ill-posed inverse problem which is why a sophisticated algorithm is necessary to perform this operation. The PSF of an optical imaging system in-focus is an Airy disk which is displayed in figure 3.4 on the right side. On the left side of figure 3.4 the airy function is displayed. To simplify the deconvolution, it is common to use a gaussian function

$$g(r) = I_0 \cdot e^{-\frac{r^2}{2\sigma^2}} \tag{3.13}$$

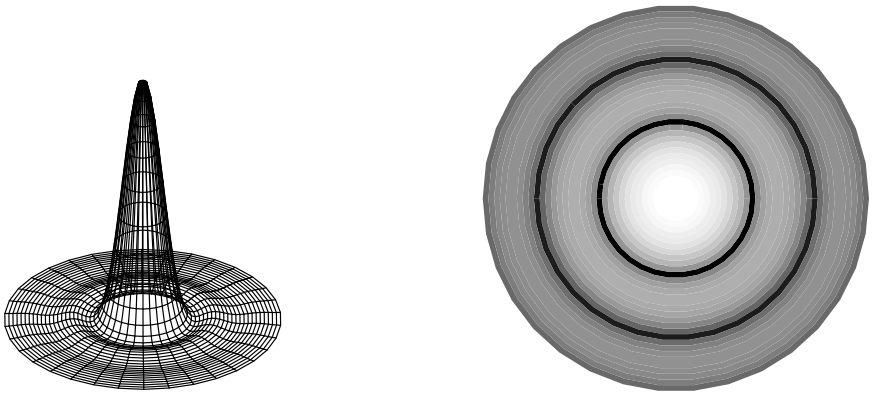

Figure 3.4: Airy Function (*left*) and corresponding Airy Disc (*right*)

which approximates the airy function fairly well as can be seen in figure 3.5. The Airy disk has its first zero, in normalized units, at

$$r = 3.8317 \tag{3.14}$$

while the standard deviation, σ, of the Gaussian approximation (also in normalized units) is given by:

$$\sigma = 1.4137 \tag{3.15}$$

This approximation ensures that the integral over the Gaussian approximation is the same as the integral over the Airy function.

Although the Airy disk theoretically extends to infinity, it can be seen from figure 3.4, that the airy function drops quite rapidly with 83.8% of its total power being contained within its first ring [43]. If the pixel size is large, relative to the Airy disk, the resolution is not diffraction limited but limited by the pixel size of the array. It is therefore not possible to gain resolution by deconvolving with the PSF.

It is worth noting, however, that the reconstructed signal from the algorithm described above is sampled a lot finer than the original image. Thus it is theoretically possible to perform the deconvolution with the PSF after the steps described in Section 3.2, even for a pixel size to PSF ratio where it was not possible in the first place. The results of deconvolving with a PSF are shown in Section 4.2.1.

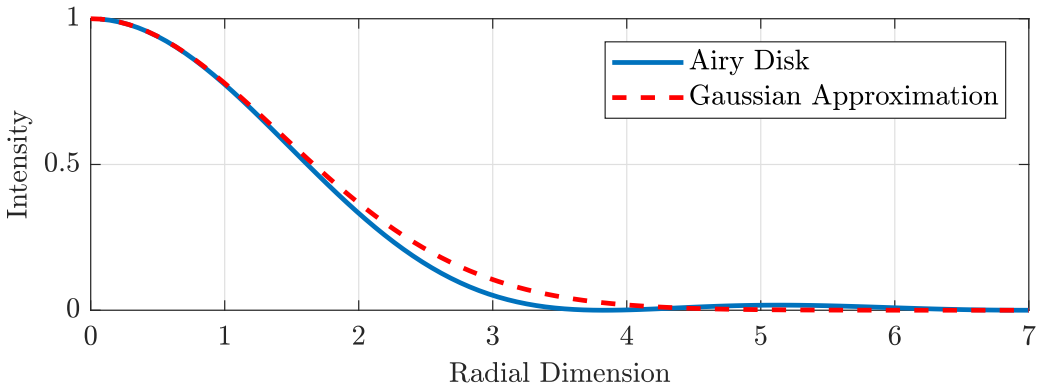

Figure 3.5: Gaussian Approximation to the Airy Function,

3.3.2 The effect of Noise on the Signal

During the imaging process, there are several different sources of noise: thermal noise, Gaussian noise and shot noise [44]. In this master's thesis, two different approaches to dealing with the noisy signal were compared. After the pre-processing steps described in Section 3.2, but before the deconvolution with the square function, the signal was either filtered with a moving average filter equal to the number of frames added to create the 1D-signal ($\rightarrow n_{step,max}$ in the ideal case) which was done with the built in MATLAB function *filtfilt* which does not introduce any phase shifts to the signal or the fourier transform of the signal was multiplied with a box function thereby filtering out any frequency components higher than the size of the box function. The two methods are compared in Section 4.2.2.

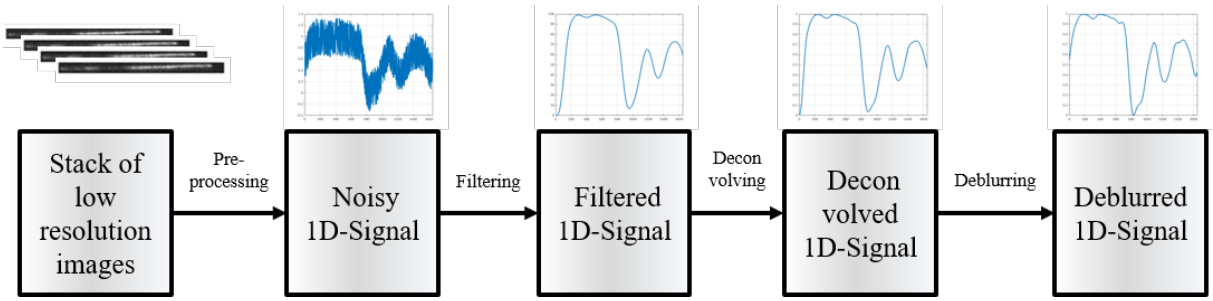

Figure 3.6: Flow Chart of the algorithm derived in Chapter 3.

Figure 3.6 summarizes the different steps of the algorithm explained above. First, the pre-processing steps described in Section 3.1 are applied to the stack of low resolution images by which the 1D signal, y_{1D}, is created, which is afterwards filtered with a moving average filter. This filtered signal is then deconvolved with a square function to restore the original barcode

and will be called f_{blur} throughout the rest of this thesis. To fully reconstruct the original signal, a deblurring step ($\rightarrow$ deconvolving with the PSF) will be performed at the end and this created signal will be denoted by f_{recon}. The feasibility of these different steps is discussed in the following sections.

Chapter 4

Results

In this section, the results of the theoretical simulation, both with and without noise and blur are presented, followed by the results of the experimental work. The chapter will start with the results of the theoretical simulation without noise and blur. These results give a good estimate of the inherent limits of the *Lucy-Richardson* deconvolution algorithm. After this, the influence of noise and blur are addressed which gives further insight into the practical constraints of the algorithm. The third section shows the performance of the algorithm on experimentally-obtained data. The algorithm was evaluated by taking a degraded signal, letting the algorithm run over it, and afterwards evaluating if the algorithm improved the resolution of the reconstructed signal compared to the degraded signal. To evaluate the quality of the reconstruction the *root-mean-square-error (RMSE)* was chosen. It is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(DNA_n - Signal_n)^2}{N}} \tag{4.1}$$

where *DNA* denotes the reference signal and *Signal* denotes the reconstructed signal. For the theoretical simulation, it was straight forward to calculate the *RMSE* because the reference was given. For the experimental results, this was harder since the actual, analog barcode signal was unknown. The way that this issue was circumvented is explained in Section 4.3. The RMSE was chosen because it is a good measure of the accuracy of the reconstructed signal (*Signal*) relative to the actual signal (*DNA*). To effectively compare different RMSE's, it was necessary

to normalize both *Signal* and *DNA* in the following manner:

$$Signal = Signal - min(Signal) \tag{4.2}$$

$$Signal = \frac{Signal}{max(Signal)} \tag{4.3}$$

$$DNA = DNA - min(DNA) \tag{4.4}$$

$$DNA = \frac{DNA}{max(DNA)} \tag{4.5}$$

Dividing the squared errors by the number of sample points in equation 4.1 ensures that two equally good fits, differing only in the number of sampling points, will have the same RMSE value. A perfect fit would lead to a RMSE value of 0 and an arbitrarily bad fit would lead to an arbitrarily large RMSE. Figure 4.1 shows three reconstructed signals with three different RMSE values. The green line is equal to the mean of the original data shown in red. The RMSE between the original signal and this line is equal to 0.29, which can be regarded as a boundary for when all information is lost.

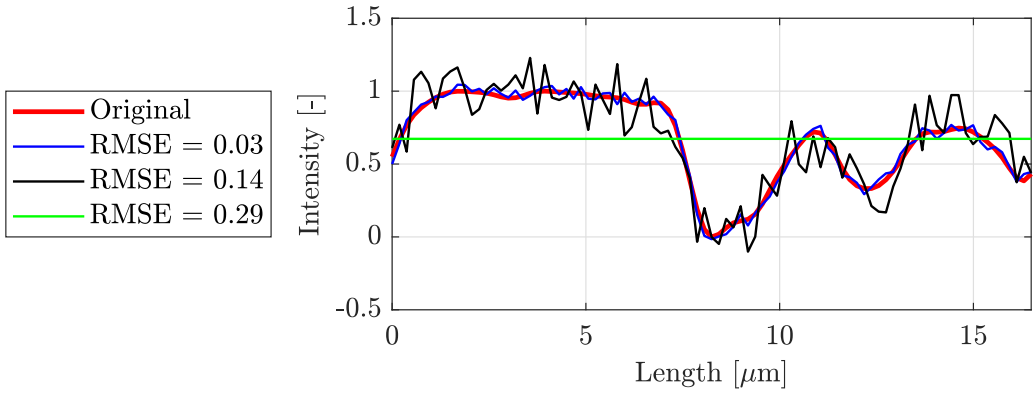

Figure 4.1: Three signals with different Root Mean Square Errors to give a magnitude estimate for the RMSE values in the following section.

4.1 Theoretical Simulation without Noise or Blur

This section shows how good the reconstruction of a signal is, when it is first convolved with a square function and afterwards deconvolved with the same square function again. The reconstructed signal will be degraded to some degree since deconvolution is an ill-posed inverse problem (see Section 3.2.1). This is done to evaluate the general performance of the *Lucy-Richardson* deconvolution algorithm presented in Section 3.2.1. It is worth noting that the performance of the reconstruction depends on the step-size l_{step} chosen. Figure 4.2 shows the

RMSE's for step sizes with a low *lcm* with the pixel sizes analyzed in the top left, and the RMSE's for step-sizes with a high *lcm* with the pixel sizes analyzed in the bottom left. The reconstruction with the highest RMSE for either is shown to the right of the RMSE plots.

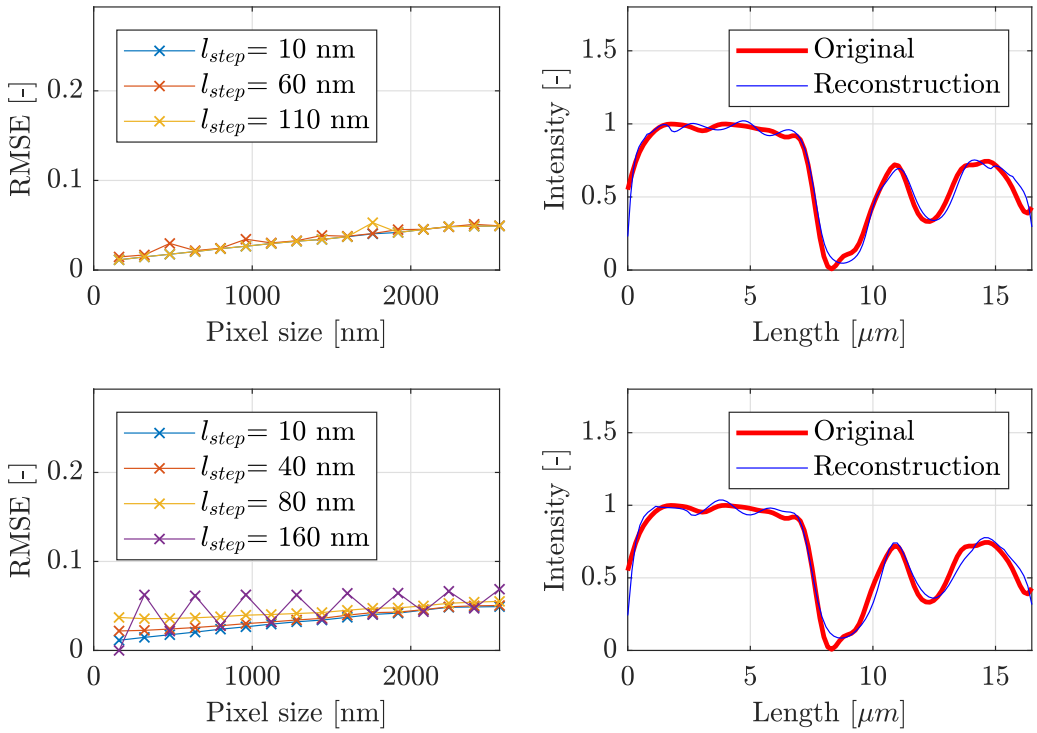

Figure 4.2: *Top Left:* RMSE for step-sizes with a low *lcm* with the pixel size and their worst reconstruction (*Top Right*). *Bottom Left*: RMSE for step-sizes with a high *lcm* with the pixel size and their worst reconstruction (*Bottom Right*).

A high *lcm* will lead to a low $n_{step,max}$ and a high l_{sample}, which means that the barcoded DNA signal is more coarsely sampled than for a lower *lcm*. It is important to keep this fact in mind when conducting experiments where the DNA molecule is moved between adjacent frames with a platform and the step-size can be controlled precisely. It is, however, worth noting that even for the poor step-size choices the *deconvlucy* algorithm is performing well. Even for a pixel size of 2.56 µm the RMSE stayed below 0.1 and the reconstruction is close to the original (as can be seen in the bottom right of figure 4.2).

In the next section, the influence of noise and blur on the simulation will be investigated which adds more parameters to the simulation. Since the performance of the algorithm is quite independent of the step-size (for a reasonably chosen step-size) the step-size will be kept constant at a value of 50 nm in the following section.

4.2 Theoretical Simulation with Noise and Blur

In this section, the influence of blur due to the PSF of the imaging system and the influence of noise on the performance of the algorithm are analyzed. The section starts by analyzing only blur (subsection 4.2.1) and only noise (subsection 4.2.2), and then looking at the combined effect of the two (subsection 4.2.3).

4.2.1 The Influence of Blur

As mentioned at the end of Section 4.1, the step-size was kept constant at 50 nm throughout this section. In this section, the original barcoded DNA signal $f(x)$ was first blurred with a gaussian function (as described in Section 3.3.1) and afterwards convolved with a square-function in the same manner as described in section 4.1. Afterwards, this signal was deconvolved with a square function, thereby reconstructing the blurred barcode. This reconstruction was denoted by $f_{blur}(x)$. Afterwards, $f_{blur}(x)$ was deblurred with the gaussian function ($\rightarrow$ deconvolved with the gaussian function) and the reconstructed barcode was denoted as $f_{recon}(x)$.

Figure 4.3 and Figure 4.4 show on the left side the RMSE between the original barcode $f(x)$ and $f_{blur}(x)$ on the one hand and $f_{recon}(x)$ on the other. The RMSE is shown for different sizes of the PSF (different standard deviations $\rightarrow std$ of the gaussian function). On the right side of both figure 4.3 and figure 4.4 is the original signal and the worst reconstruction of f_{blur} and f_{recon} shown. By comparing figure 4.3 with figure 4.4, it can be seen that deblurring the deconvolved

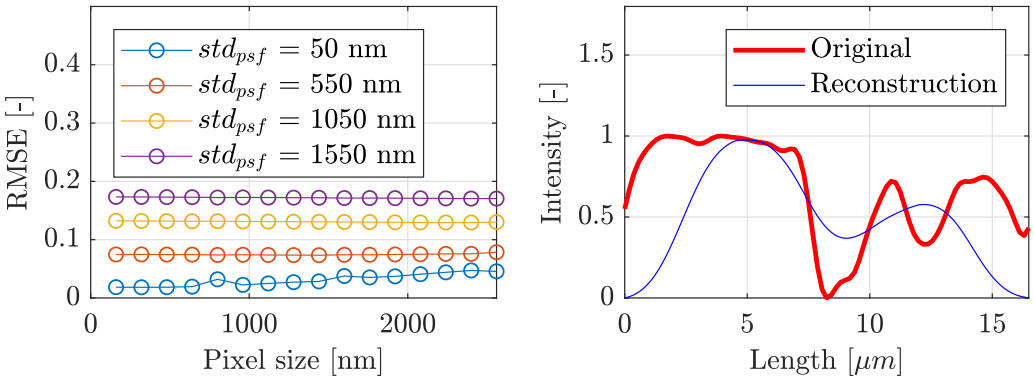

Figure 4.3: RMSE of the Deconvolved Signal before deblurring and the Original Barcode ($l_{step} = 50nm$)

signal can indeed decrease the RMSE and increase the resolution between the reconstruction

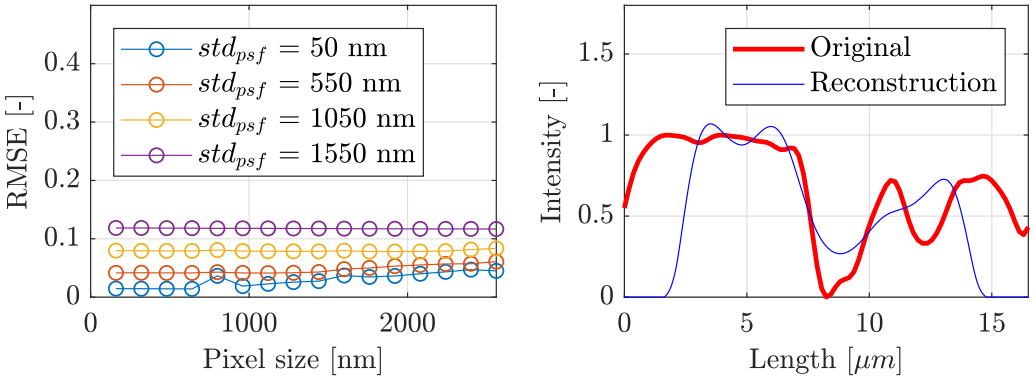

Figure 4.4: RMSE of the deblurred signal and the original barcode ($l_{step} = 50nm$).

and the original signal. The plots of the worst reconstruction and the original on the right sides of figure 4.3 and figure 4.4 confirm this: the deblurred signal f_{recon} in figure 4.4 is closer to the original than the signal before the deblurring f_{blur} in figure 4.3. It is important to note that these results were obtained with the exact knowledge of the PSF. For the implementation with experimental data, it is important to check the tolerances of the deblurring step (e.g. how much can the std deviate from the actual PSF before the resolution gets worse than $f_{blur}(x)$). Furthermore, since the actual PSF is an Airy function (see Section 3.3.1), there will be some error due to the gaussian approximation. Figure 4.5 shows the RMSE for a few deblurred signals $f_{recon}(x)$ where the gaussian used for the deblurring was smaller or larger than the actual Airy-function PSF. The reconstructed signal without the deblurring step $f_{blur}(x)$ is shown in red as a reference. The std of the actual PSF was kept constant at 550 nm in figure 4.5 and the Airy function for the PSF was calculated according to Section 3.3.1.

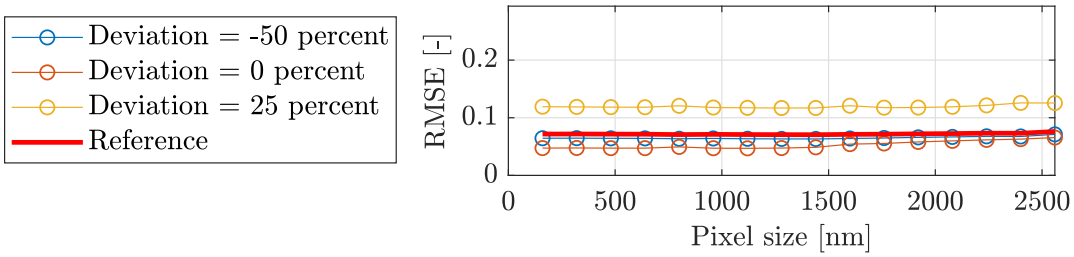

Figure 4.5: Deblur with a Gaussian that is larger/smaller than the actual PSF ($\sigma_{actual} = 550nm$ & $l_{step} = 50nm$)

Figure 4.5 shows that the resolution of the signal can still be increased if the PSF is estimated exactly right or slightly too small. If the PSF is estimated too large, however, the RMSE is higher than without performing the deblurring step. This can be seen by the yellow line, where

the std was estimated 25 % to large.

4.2.2 The Influence of Noise

As mentioned in Section 3.3.2, there are many sources of noise during the imaging process. In this section, two methods for filtering out noise are analyzed. On the one hand, a moving average filter, and on the other, a low-pass Fourier filter. The size of the moving average filter was equal to the number of frames used to create the 1D-signal (which was $n_{step,max}$ for the theoretical simulation here).

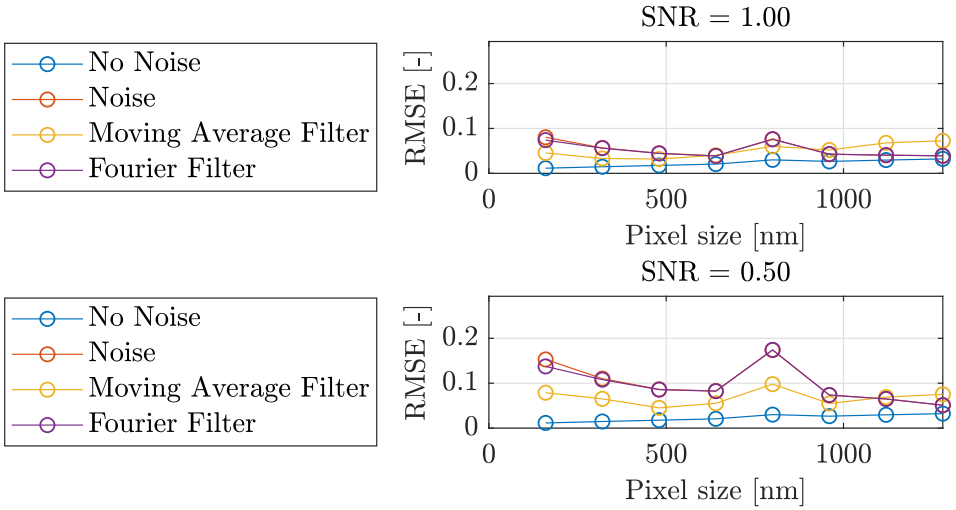

Figure 4.6: RMSE of the noisy signal and the signal filtered with a moving-average-filter and a Fourier filter. As a reference, the RMSE of the signal without noise. On top: SNR = 1 and on the bottom: SNR = 0.5 ($l_{step} = 50nm$)

The low-pass Fourier filter worked in the following manner: first the signal was Fourier transformed, then the Fourier transform of the signal was multiplied with a box-function with a size of

$$f_{box} = \frac{1}{cmr \cdot l_{basepair}} \tag{4.6}$$

thereby only keeping frequency components below this limit. As described in Section 3.2, denaturation mapping has an inherent resolution limit due to the *cooperatively melting regions* in the DNA molecule. As described in Section 3.2, this number is a few hundred base pairs. For this simulation, it was chosen to be $cmr = 300$.

Figure 4.6 shows the RMSE for two different *SNR* (Signal-to-Noise) ratio, SNR = 1 on top and

SNR = 0.5 on the bottom. It can be seen that, in general, filtering with a mean-average-filter decreases the RMSE between the original and the reconstruction whereas filtering the signal with the Fourier filter does not. Because of this, the moving-average-filter was chosen for the data processing described in Section 4.3. The Fourier filtering with a higher *cmr* value of 2000 showed an improvement of the RMSE and is described in Section 4.3.2.

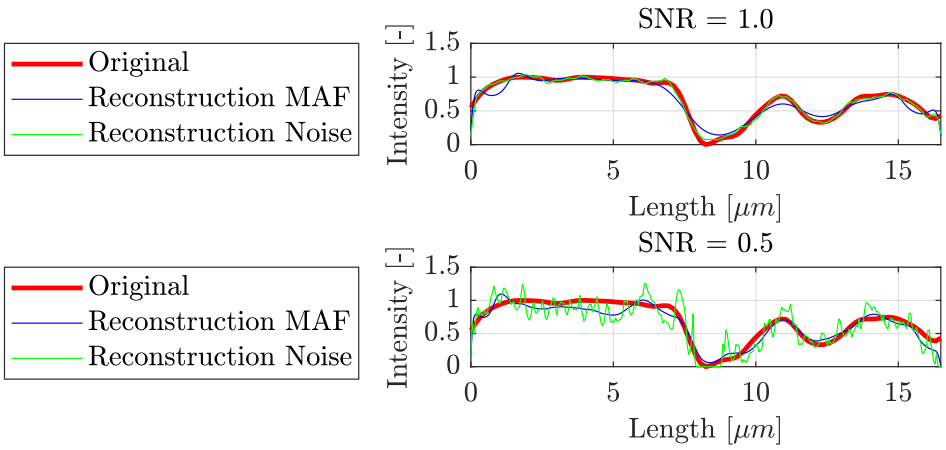

Figure 4.7: *Top*: Highest RMSE fit of the noisy signal (green) and the signal smoothed with a moving average filter (black) for a SNR = 1. *Bottom*: Highest RMSE fit of the noisy signal (green) and the signal smoothed with a moving average filter (black) for a SNR = 0.5

4.2.3 The Influence of Blur and Noise

In this section, the combined influence of blur and noise is analyzed. In order to do this the original barcoded DNA signal $f(x)$ was first blurred with a PSF of a size of 550 nm and afterwards, noise with SNR = 0.5 was added to the signal. This blurred, noisy signal was then convolved with a square-function to simulate the imaging process. After that, the signal was restored. This was done by first filtering out the noise with a moving-average-filter (described in Section 4.2.2) and then by deconvolving the filtered barcode with a square function to reverse the imaging process. As in Section 4.2.1, this signal was denoted by $f_{blur}(x)$. After all of that, as a final step, the signal was deblurred with a gaussian-function. This was done to see if it was still possible to achieve a gain in resolution by deblurring after the filtering and square-deconvolution steps. Again, as in Section 4.2.1, this signal was denoted by $f_{recon}(x)$. Figure 4.8 shows the RMSE between $f_{blur}(x)$ and $f(x)$ on the top and f_{recon} and $f(x)$ on the bottom.

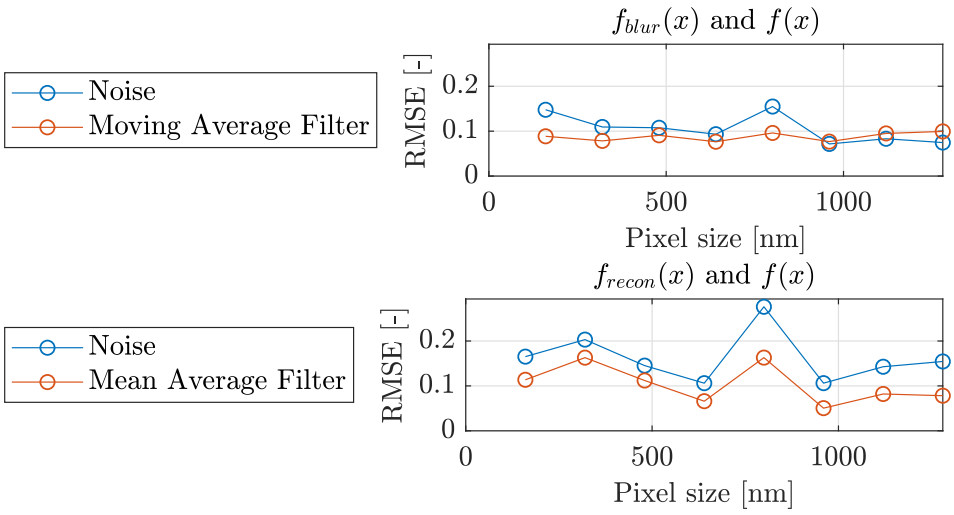

Figure 4.8: RMSE for the signal before and after deblurring ($SNR = 0.5$ and $l_{step} = 50nm$)

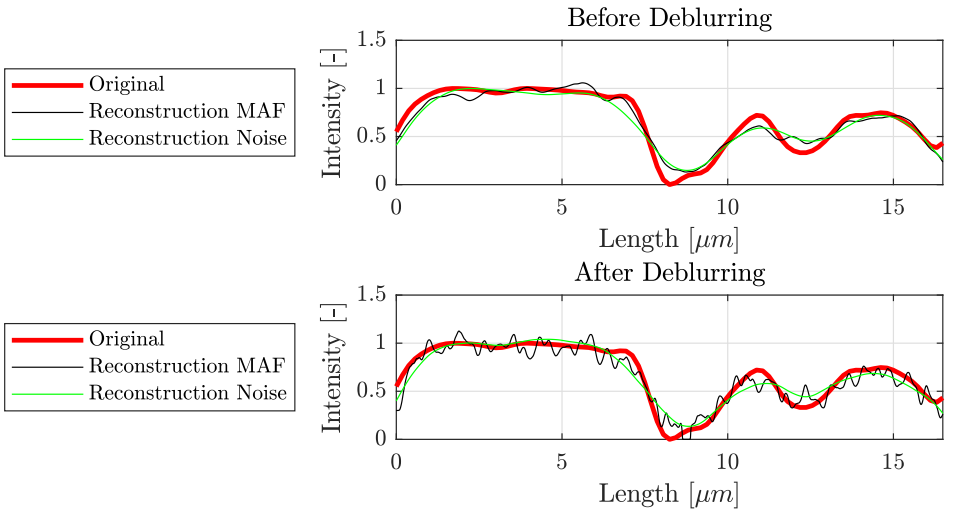

Figure 4.9: *Top*: Highest RMSE fit of the noisy signal (green) and the signal smoothed with a moving average filter (black) before the deblurring step. *Bottom*: Highest RMSE fit of the noisy signal (green) and the signal smoothed with a moving average filter (black) after the deblurring step. ($SNR = 0.5$ and $l_{step} = 50nm$)

Figure 4.8 shows that the deblurring step is detrimental to the resolution of the restored barcode, even though it is theoretically possible to increase the quality of the reconstruction. The reason for this degradation of the restoration must be because of the nature of the deconvolution problem. Figure 4.9 shows the worst reconstruction of moving-averaged signals in black, and the worst reconstructions of the noisy signal in green. The top half of figure 4.9 shows the signals before the deblurring step and the bottom half shows the signals after the deblurring

step.

4.3 Experimental Results

This section will present the experimental results obtained during this project. First, the performance of the algorithm on images of fluorescent beads is shown in Section 4.3.1. Then, the performance of the algorithm on DNA molecules prepared with denaturation mapping is shown in Section 4.3.2. An important concept for the following sections is the concept of binning pixels together, which means combining several pixels into one. This will increase the pixel size by the binning factor and decrease the resolution concomitantly. See Figure 4.10.

As already mentioned in the beginning of chapter 4, it is not trivial to define a reference

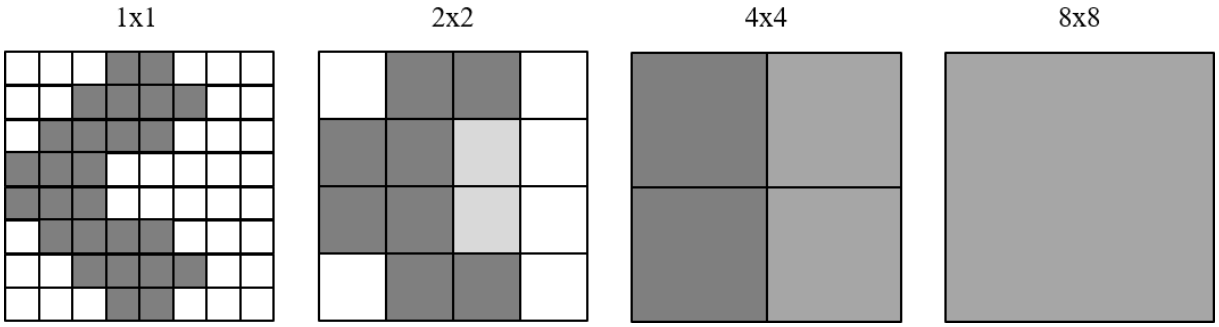

Figure 4.10: Binning scheme showing how the pixel size increases from an unbinned image (1x1), up to a binning of 8x8.

when using the algorithm on experimental data. For the next sections, the following scheme was used: First, the unbinned image-stack was processed by the algorithm. This created the reference, *DNA*, in equation 4.1. Afterwards, the binning of the image stack was increased, thereby simulating an imaging system with lower resolution. Next, the image-stack was again processed by the algorithm. This created the reconstruction, *Signal*, in equation 4.1. Then, the RMSE between the unbinned signal and the signal obtained from the binned stack of images was calculated. It is worth noting that the number of frames required for the algorithm differed depending on the binning factor. For example, for a step-size of 40 nm and an (unbinned) pixel size of 160 nm, the number of frames required by the algorithm, $n_{step,max}$, is 4. This number changes however when the binning is increased from 8 (2x2) to 12 (3x3), 16 (4x4), 20 (5x5), 24 (6x6), 28 (7x7) up to 32 (8x8). This is because the binning increases the pixel-size (see formula 3.4).

To evaluate the performance of the algorithm, the following reference was used. A single frame was taken and a linear signal was created by summing over the dimension perpendicular to the movement of the FOV (this is equivalent to only performing the first step of the pre-processing as described in Section 3.2). For the reference, the binning was also increased and then the RMSE was calculated between the binned signals and the unbinned signal.

4.3.1 Analysis of Nanometer-sized Beads

The feasibility of the algorithm was first evaluated with an image stack of fluorescent beads with a diameter of 60 nm. These beads have a few advantages over DNA molecules prepared with

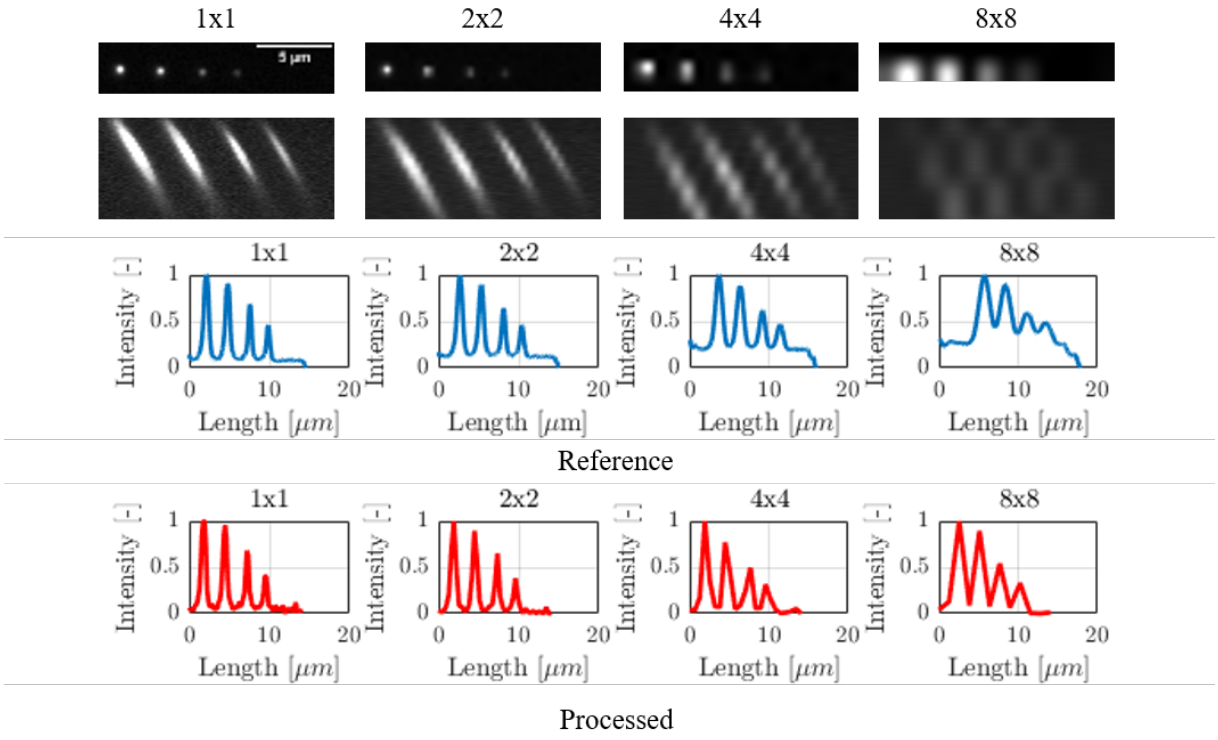

Figure 4.11: Images of the beads for different binnings (top), their reference signal (middle), processed signal (bottom).

denaturation mapping: their signal has a much higher signal to noise ratio than the one from DNA molecules; their signal is more straightforward to interpret than the barcode signal displayed by a DNA molecule; and most importantly, they show basically no photobleaching. As a proof of principle, four beads were imaged in the following manner: 32 Frames were taken with a step-size of 40 nm between them, 64 frames with a step-size of 60 nm between them and 16 frames with a step-size of 80 nm between them. These frames were then processed as described

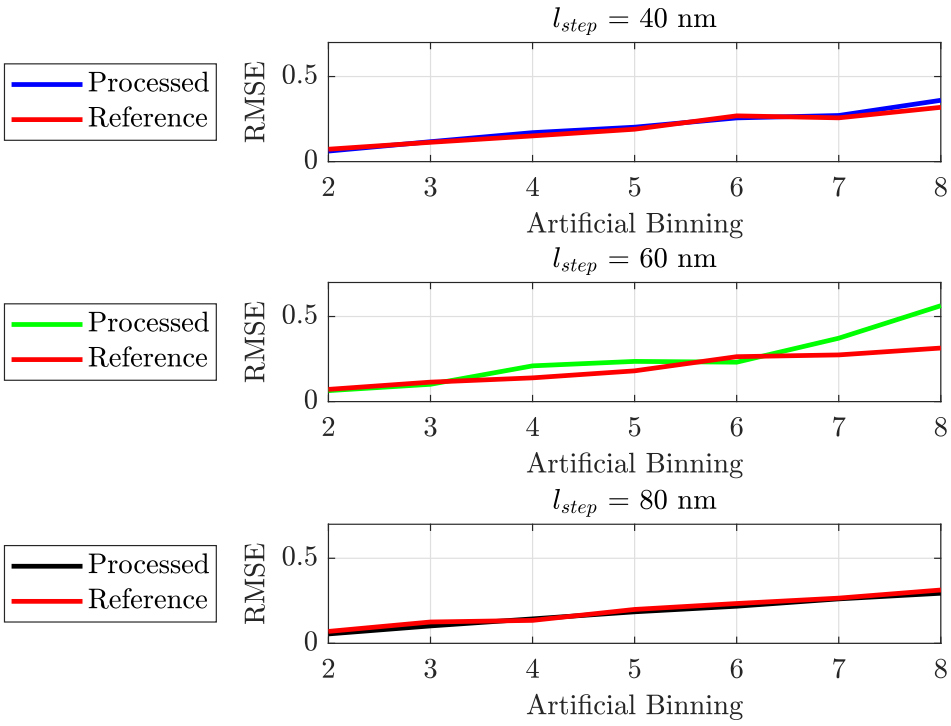

Figure 4.12: RMSE of the processed signal and the reference signal for different binnings. Step size 40 nm (top) 60 nm (middle) and 80 nm (top).

in Section 3.2. Afterwards, the pixels were binned together, starting from 2x2 all the way up to 8x8, and the processing was repeated. Finally, the RMSE between these binned signals and the unbinned signal was calculated and can be seen in figure 4.12. Figure 4.12 shows that the RMSE was more or less the same for the processed signal and the reference. Figure 4.11 shows the four beads, their reference signal and their processed signal.

4.3.2 Analysis of DNA molecules

During this master's project 65 DNA barcode signals were analyzed: 31 barcodes with a step size of 40 nm, 12 with a step size of 60 nm and 22 barcodes with a step size of 80 nm. The barcodes signals were processed in the same manner as described above. Figure 4.13 shows the average RMSE of all the barcodes for the 40 nm, 60 nm and 80 nm measurements on the left side.

Figure 4.14 shows the images of three different barcoded DNA molecules, their processed signals, and their Fourier filtered processed signals. The processed signal is still noisy, containing

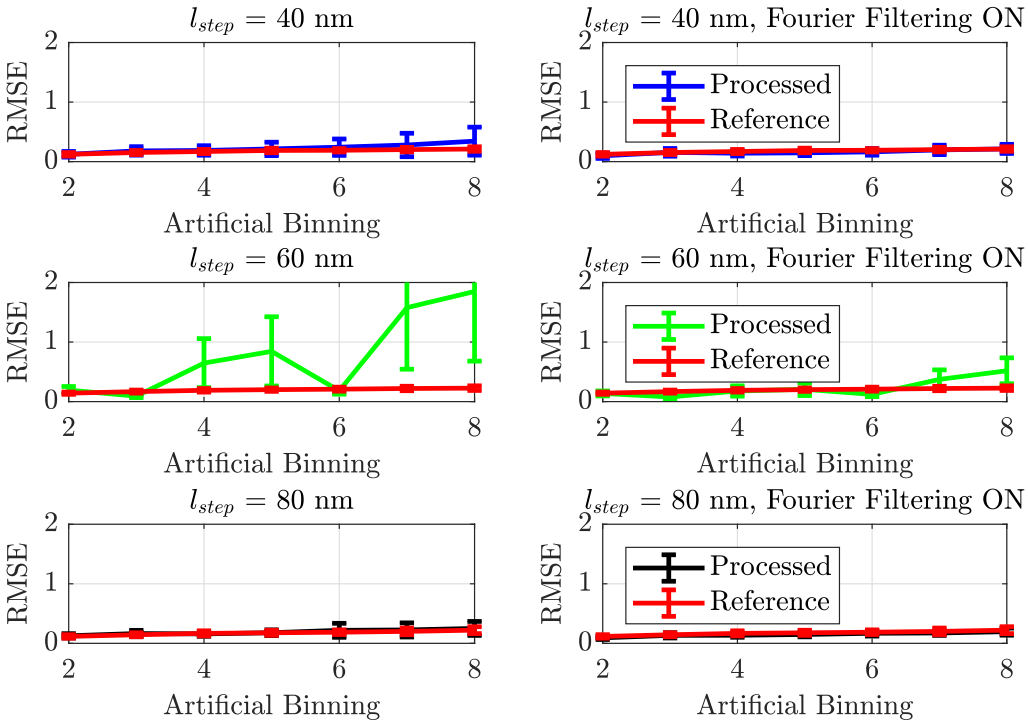

Figure 4.13: RMSE over artificial binning for the 40 nm step-size data (top, 31 barcodes), 60 nm step-size data (middle, 12 barcodes), and the 80 nm step size data (bottom, 22 barcodes). Left side: processed without Fourier-filtering, right side: processed with Fourier-filtering.

high frequency components that contain no information about the DNA molecule analyzed and which the Fourier filtering takes care of. For these images, all frequencies below 1.4706 µm^{-1} have been filtered out. If the Fourier filtering is applied for all the different binnings after the processing, the RMSE of the processed signal can be lowered. This can be seen on the right side of figure 4.13. The cut-off frequency of 1.4706 µm^{-1} corresponds to:

$$\frac{1}{cmr \cdot l_{basepair}} = \frac{1}{2000 \cdot l_{basepair}} \tag{4.7}$$

This number has been empirically estimated and showed a good improvement of the RMSE as can be seen in figure 4.13. Figure 4.15 shows the signals of the three barcodes displayed in figure 4.14. As expected the signal degraded with higher binning. The reconstructed resolution was not as good as the theoretical simulations displayed in Section 4.1 suggested. Figure 4.16 shows the signals of figure 4.15 after Fourier filtering. It can be seen that only high frequency components are filtered out that carry no information about the analyzed molecule. This leads to a lower RMSE, as can be seen on the right side of figure 4.13.

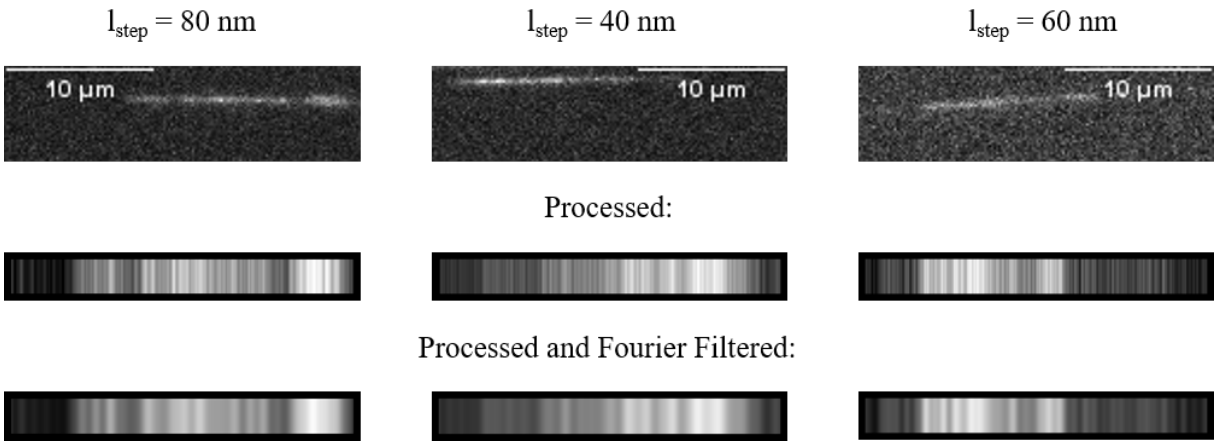

Figure 4.14: Images of three barcoded DNA molecules (top), their processed signal (middle) and their Fourier filtered signal (bottom)

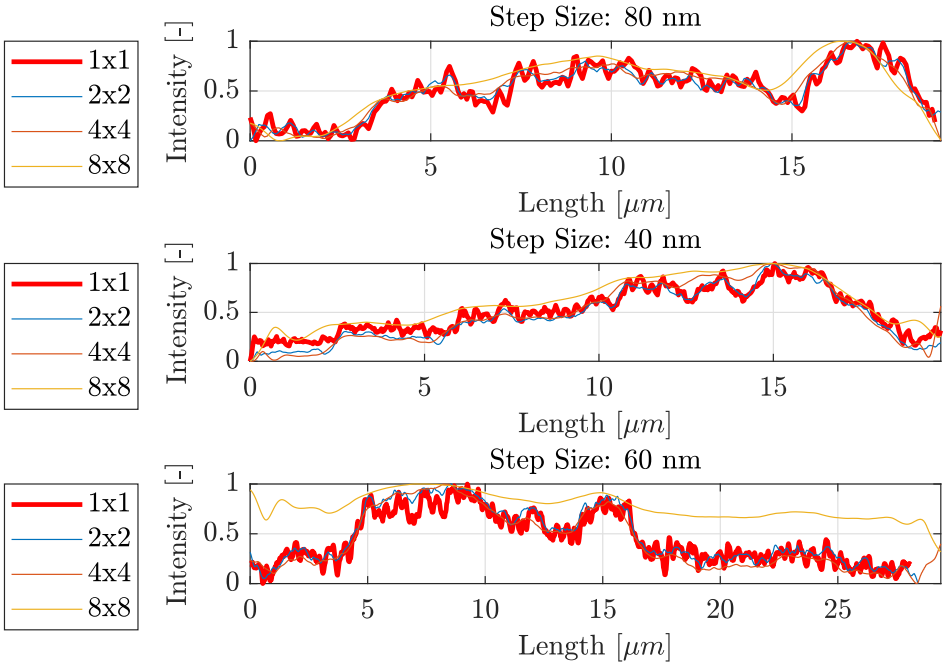

Figure 4.15: Barcode signals of the barcodes shown in figure 4.14 before the Fourier filtering step.

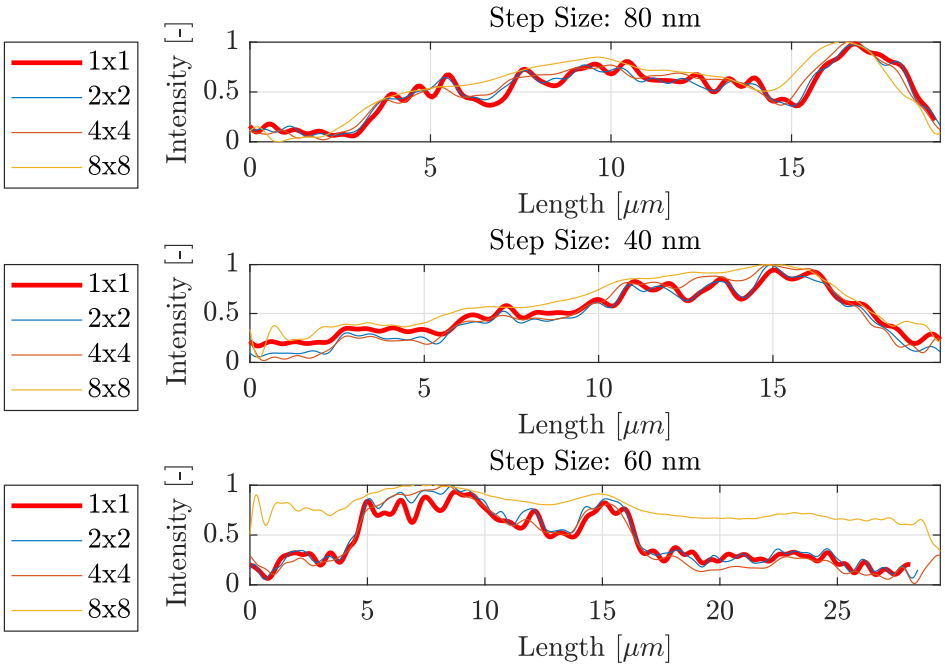

Figure 4.16: Barcode signals of the barcodes shown in figure 4.14 after the Fourier filtering step.

Chapter 5

Discussion and Outlook

In this master's thesis an algorithm for the processing of images of DNA molecules prepared with denaturation mapping was developed. The input for the algorithm was stacks of images with a controlled unidirectional shift between them. The algorithm used those images to create a one-dimensional signal with a higher resolution than the signal of a single frame. The theory behind the algorithm was developed in chapter 3 and its practical limitations have been tested on artificial data in Section 4.1. Afterwards, the algorithm was tested on experimental data which was described in Section 4.3. The algorithm was first evaluated on fluorescent beads, which, due to their simplicity, have been simpler to analyze than DNA molecules prepared with denaturation mapping. The later have been analyzed afterwards in Section 4.3.2. The algorithm has shown to increase the resolution of the barcoded DNA signal compared to the signal of a single frame for a few of the molecules. A few of the problems encountered, however, will be discussed below.

One of the major limitations of using the method presented in this thesis in a real experimental setup is *photobleaching*. Photobleaching is caused by damage to dyes, to the DNA itself and to conformational changes in the DNA and it leads to the intensity of the barcoded DNA signal getting dimmer with every frame taken. Not only that, since the signal decreases homogeneously there appear to be new features in the bleached molecule. This drastically limits the number of frames that can be taken of a single DNA molecule. Figure 5.1 shows a barcoded DNA molecule for the first frame imaged in the left half, and for the 39th frame on the right. The intensity of the signal in the 39th frame has dropped to one third of the intensity in the first frame. In theory, the resolution of the barcoded DNA signal produced by the algorithm gets better with the number of frames taken. This is why, theoretically, l_{step} should be chosen

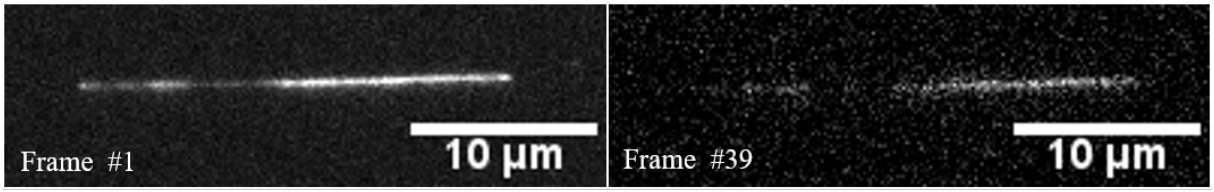

Figure 5.1: *Left:* Barcoded DNA molecule. *Right:* Photobleached barcoded DNA molecule

(in relation to l_{pixel}) so as to maximize $n_{step,max}$. However, because the intensity of the barcode signal bleaches out over time, this puts an upper bound on $n_{step,max}$ and therefore on the value of l_{step}. If the number of frames taken is smaller than $n_{step,max}$, there is still a possibility to use the algorithm. The signal can be interpolated before the deconvolution step, creating an evenly spaced signal which can then be restored by deconvolving with a square function. It is important to keep in mind though that this step can be a source of error.

In this thesis the groundwork of the algorithm has been laid out and a proof-of-principle has been established. To implement the algorithm in a smartphone microscope or similar optical device, there are still some issues that need to be addressed. The first and probably most urgent one is to find a way to ensure controlled movement of the DNA molecule over the whole imaging process and to analyze how deviations in step size will affect the reconstruction algorithm. Furthermore, the influence, or rather the feasibility, of deblurring (i.e. deconvolving with the PSF) and at what point of the algorithm it can be implemented should be further analyzed. Since

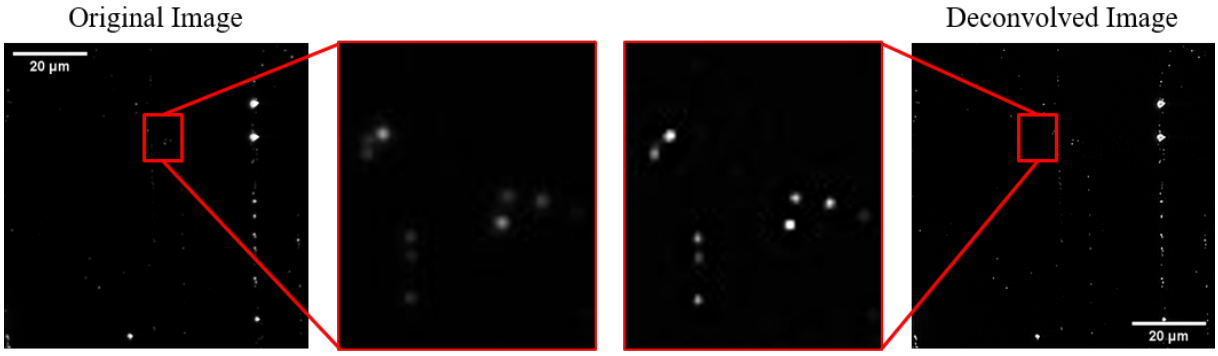

Figure 5.2: PSF estimation with images of 100 nm sized beads (→ small dots).

all experimental data during this master thesis was acquired with a state-of-the-art fluorescent microscope, the influence of blur due to the imaging system was almost negligible. An analysis with fluorescent 100 nm sized beads showed that the std of the PSF was around 300 nm. Figure 5.2 shows the result of the deblurring. There is an increase in resolution between the left and the right image in figure 5.2. It needs to be noted, however, that this was an image without bin-

ning, with a pixel size of 160 nm. If the pixel size was increased by binning two or more pixels together the deblurring step would no longer be feasible since almost the entire PSF would be contained in a single pixel. The 100 nm beads can be identified as the small dots in the image. Another step toward implementing and automating the algorithm is to combine it with a detection algorithm of the DNA molecule. In this way, it would only be necessary to move the barcoded DNA molecule through the FOV while the software independently and automatically reconstructs the high-resolution signal.

Acknowledgements

I want to thank my supervisors Prof. Jonas Tegenfeldt, Dr. Jason Beech and Oskar Ström for coming up with the idea for this project, helping me conduct the necessary experiments and constantly giving me valuable feedback. This project would not have been possible without them. Furthermore I want to thank the other members of Prof. Jonas Tegenfeldts' group, Dr. Elke Heibisch, Kushagr Punyani, Dr. Bao Ho Dang, Trung Tran and Feifei Peng. Again without their great feedback during the group meetings this thesis would not have been possible. Furthermore I want to thank Assoc.Prof. Tobias Ambjörnsson and Albertas Dvirnas. One of the goals of mine was to combine the presented algorithm with the DNA alignment algorithm presented in [45]. Unfortunately I have not been able to accomplish this but I still want to thank Tobias and Albertas for sharing their code with me and for offering their help. I also want to thank Ida Arvidsson from *Matematik LTH* for answering some of my questions around deconvolution algorithms.

Finally I want to thank my family, my friends and my office mates for constantly supporting me, especially my great friend Austin Irish for helping me with the corrections of my thesis.

Chapter 6

References

[1] M. Oheim, "High-throughput microscopy must re-invent the microscope rather than speed up its functions," *British journal of pharmacology*, vol. 152, no. 1, pp. 1–4, 2007.

[2] X. Cui, L. M. Lee, X. Heng, W. Zhong, P. W. Sternberg, D. Psaltis, and C. Yang, "Lensless high-resolution on-chip optofluidic microscopes for caenorhabditis elegans and cell imaging," *Proceedings of the National Academy of Sciences*, vol. 105, no. 31, pp. 10670–10675, 2008.

[3] X. Heng, D. Erickson, L. R. Baugh, Z. Yaqoob, P. W. Sternberg, D. Psaltis, and C. Yang, "Optofluidic microscopy—a method for implementing a high resolution optical microscope on a chip," *Lab on a Chip*, vol. 6, no. 10, pp. 1274–1276, 2006.

[4] X. Heng, X. Cui, D. W. Knapp, J. Wu, Z. Yaqoob, E. J. McDowell, D. Psaltis, and C. Yang, "Characterization of light collection through a subwavelength aperture from a point source," *Optics express*, vol. 14, no. 22, pp. 10410–10425, 2006.

[5] W. Bishara, S. O. Isikman, and A. Ozcan, "Lensfree optofluidic microscopy and tomography," *Annals of biomedical engineering*, vol. 40, no. 2, pp. 251–262, 2012.

[6] G. Zheng, S. A. Lee, S. Yang, and C. Yang, "Subpixel resolving optofluidic microscope based on super resolution algorithm," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pp. 1362–1365, IEEE, 2011.

[7] S. A. Lee, R. Leitao, G. Zheng, S. Yang, A. Rodriguez, and C. Yang, "Color capable sub-pixel resolving optofluidic microscope and its application to blood cell imaging for malaria diagnosis," *PloS one*, vol. 6, no. 10, p. e26127, 2011.

[8] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE transactions on image processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[9] B. Guo, C. Lei, Y. Wu, H. Kobayashi, T. Ito, Y. Yalikun, S. Lee, A. Isozaki, M. Li, Y. Jiang, *et al.*, "Optofluidic time-stretch quantitative phase microscopy," *Methods*, 2017.

[10] P. Memmolo, L. Miccio, F. Merola, O. Gennari, P. A. Netti, and P. Ferraro, "3d morphometry of red blood cells by digital holography," *Cytometry part A*, vol. 85, no. 12, pp. 1030–1036, 2014.

[11] R. Wang, Z. Wang, L. Millet, M. U. Gillette, A. Levine, and G. Popescu, "Dispersion-relation phase spectroscopy of intracellular transport," *Optics Express*, vol. 19, no. 21, pp. 20571–20579, 2011.

[12] H. Ding, Z. Wang, F. Nguyen, S. A. Boppart, and G. Popescu, "Fourier transform light scattering of inhomogeneous and dynamic structures," *Physical review letters*, vol. 101, no. 23, p. 238102, 2008.

[13] B. Guo, C. Lei, H. Kobayashi, T. Ito, Y. Yalikun, Y. Jiang, Y. Tanaka, Y. Ozeki, and K. Goda, "High-throughput, label-free, single-cell, microalgal lipid screening by machine-learning-equipped optofluidic time-stretch quantitative phase microscopy," *Cytometry Part A*, vol. 91, no. 5, pp. 494–502, 2017.

[14] A. K. Lau, T. T. Wong, K. K. Ho, M. Y. Tang, A. C. Chan, X. Wei, E. Y. Lam, H. C. Shum, K. K. Wong, and K. K. Tsia, "Interferometric time-stretch microscopy for ultrafast quantitative cellular and tissue imaging at 1 μm," *Journal of biomedical optics*, vol. 19, no. 7, p. 076001, 2014.

[15] H. Ceylan Koydemir and A. Ozcan, "Smartphones democratize advanced biomedical instruments and foster innovation," *Clinical Pharmacology & Therapeutics*, vol. 104, no. 1, pp. 38–41, 2018.

[16] D. Tseng, O. Mudanyali, C. Oztoprak, S. O. Isikman, I. Sencan, O. Yaglidere, and A. Ozcan, "Lensfree microscopy on a cellphone," *Lab on a Chip*, vol. 10, no. 14, pp. 1787–1792, 2010.

[17] S. A. Lee and C. Yang, "A smartphone-based chip-scale microscope using ambient illumination," *Lab on a Chip*, vol. 14, no. 16, pp. 3056–3063, 2014.

[18] S. Kheireddine, A. S. Perumal, Z. J. Smith, D. V. Nicolau, and S. Wachsmann-Hogiu, "Dual-phone illumination-imaging system for high resolution and large field of view multi-modal microscopy," *Lab on a Chip*, 2019.

[19] Q. Wei, H. Qi, W. Luo, D. Tseng, S. J. Ki, Z. Wan, Z. Göröcs, L. A. Bentolila, T.-T. Wu, R. Sun, *et al.*, "Fluorescent imaging of single nanoparticles and viruses on a smart phone," *ACS nano*, vol. 7, no. 10, pp. 9147–9155, 2013.

[20] H. C. Koydemir, Z. Gorocs, D. Tseng, B. Cortazar, S. Feng, R. Y. L. Chan, J. Burbano, E. McLeod, and A. Ozcan, "Rapid imaging, detection and quantification of giardia lamblia cysts using mobile-phone based fluorescent microscopy and machine learning," *Lab on a chip*, vol. 15, no. 5, pp. 1284–1293, 2015.

[21] Q. Wei, G. Acuna, S. Kim, C. Vietz, D. Tseng, J. Chae, D. Shir, W. Luo, P. Tinnefeld, and A. Ozcan, "Plasmonics enhanced smartphone fluorescence microscopy," *Scientific reports*, vol. 7, no. 1, p. 2124, 2017.

[22] L. K. Nyberg, F. Persson, J. Berg, J. Bergström, E. Fransson, L. Olsson, M. Persson, A. Stålnacke, J. Wigenius, J. O. Tegenfeldt, *et al.*, "A single-step competitive binding assay for mapping of single dna molecules," *Biochemical and biophysical research communications*, vol. 417, no. 1, pp. 404–408, 2012.

[23] W. Reisner, N. B. Larsen, A. Silahtaroglu, A. Kristensen, N. Tommerup, J. O. Tegenfeldt, and H. Flyvbjerg, "Single-molecule denaturation mapping of dna in nanofluidic channels," *Proceedings of the National Academy of Sciences*, vol. 107, no. 30, pp. 13294–13299, 2010.

[24] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, "Base-stacking and base-pairing contributions into thermal stability of the dna double helix," *Nucleic acids research*, vol. 34, no. 2, pp. 564–574, 2006.

[25] C. Sadhu, S. Dutta, and K. Gopinathan, "Influence of formamide on the thermal stability of dna," *Journal of Biosciences*, vol. 6, no. 6, pp. 817–821, 1984.

[26] V. Müller and F. Westerlund, "Optical dna mapping in nanofluidic devices: principles and applications," *Lab on a Chip*, vol. 17, no. 4, pp. 579–590, 2017.

[27] H. S. Rye, S. Yue, D. E. Wemmer, M. A. Quesada, R. P. Haugland, R. A. Mathies, and A. N. Glazer, "Stable fluorescent complexes of double-stranded dna with bis-intercalating asymmetric cyanine dyes: properties and applications," *Nucleic Acids Research*, vol. 20, no. 11, pp. 2803–2812, 1992.

[28] T. Fisher, "Fluorescence spectraviewer." `https://www.thermofisher.com/se/en/home/life-science/cell-analysis/labeling-chemistry/fluorescence-spectraviewer.html`, April 2019.

[29] J. Deen, W. Sempels, R. De Dier, J. Vermant, P. Dedecker, J. Hofkens, and R. K. Neely, "Combing of genomic dna from droplets containing picograms of material," *ACS nano*, vol. 9, no. 1, pp. 809–816, 2015.

[30] W. Reisner, J. P. Beech, N. B. Larsen, H. Flyvbjerg, A. Kristensen, and J. O. Tegenfeldt, "Nanoconfinement-enhanced conformational response of single dna molecules to changes in ionic environment," *Physical review letters*, vol. 99, no. 5, p. 058302, 2007.

[31] F. Persson and J. O. Tegenfeldt, "Dna in nanochannels—directly visualizing genomic information," *Chemical Society Reviews*, vol. 39, no. 3, pp. 985–999, 2010.

[32] J. Marmur, "A procedure for the isolation of deoxyribonucleic acid from micro-organisms," *Journal of molecular biology*, vol. 3, no. 2, pp. 208–IN1, 1961.

[33] F. Li and A. Nathan, *CCD image sensors in deep-ultraviolet: degradation behavior and damage mechanisms*. Springer Science & Business Media, 2006.

[34] L. Meng, "An intensified emccd camera for low energy gamma ray imaging applications," *IEEE transactions on nuclear science*, vol. 53, no. 4, pp. 2376–2384, 2006.

[35] S. K. Madan, B. Bhaumik, and J. Vasi, "Experimental observation of avalanche multiplication in charge-coupled devices," *IEEE Transactions on Electron Devices*, vol. 30, no. 6, pp. 694–699, 1983.

[36] M. S. Robbins and B. J. Hadwen, "The noise performance of electron multiplying charge-coupled devices," *IEEE transactions on Electron Devices*, vol. 50, no. 5, pp. 1227–1232, 2003.

[37] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the I.R.E*, vol. 37, no. 1, pp. 10–21, 1949.

[38] H. D. Luke, "The origins of the sampling theorem," *IEEE Communications magazine*, vol. 37, no. 4, pp. 106–108, 1999.

[39] W. H. Richardson, "Bayesian-based iterative method of image restoration," *JOSA*, vol. 62, no. 1, pp. 55–59, 1972.

[40] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The astronomical journal*, vol. 79, p. 745, 1974.

[41] D. S. Biggs and M. Andrews, "Acceleration of iterative image restoration algorithms," *Applied optics*, vol. 36, no. 8, pp. 1766–1775, 1997.

[42] R. Liu and J. Jia, "Reducing boundary artifacts in image deconvolution," in *2008 15th IEEE International Conference on Image Processing*, pp. 505–508, IEEE, 2008.

[43] M. Born and E. Wolf, "Principles of optics pergamon press," *Chap*, vol. 3, no. 1, p. 118, 1980.

[44] A. K. Boyat and B. K. Joshi, "A review paper: Noise models in digital image processing," *arXiv preprint arXiv:1505.03489*, 2015.

[45] C. Noble, A. N. Nilsson, C. Freitag, J. P. Beech, J. O. Tegenfeldt, and T. Ambjörnsson, "A fast and scalable kymograph alignment algorithm for nanochannel-based optical dna mappings," *PloS one*, vol. 10, no. 4, p. e0121905, 2015.