



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Unterschrift des Betreuers

DIPLOMARBEIT

Influence of spatial resolution and single cell identification on bacterial cell cycle observations

Ausgeführt am Institut für
Applied Physics, Biophysics Group
der Technischen Universität Wien

unter der Anleitung von Prof. Dr. Gerhard Schütz

durch

Felix Schaber
Strohmayergasse 7/10, 1060 Wien

Datum

Unterschrift (Student)

Influence of spatial resolution and single cell identification on bacteria cell cycle observations

Felix Schaber

December 4, 2017

Abstract

Accurate description of bacterial dynamics is an important factor for understanding the bacterial cell cycle. Traditionally direct observation of cell cycle dynamics using light microscopy has been restricted by the diffraction limit. In recent years a number of super-resolution techniques have been developed allowing to overcome these limitations. However, time-lapse imaging of single cells is challenging due to photobleaching from prolonged illumination and cell movement between images. In addition, some dyes used in these techniques require the bacteria to be fixed prior to microscopy. This thesis explores the possibilities to reconstruct a pseudo-temporal trajectory from static images of cell populations and compares them to lower resolution single cell time lapse microscopy. Results indicate that pseudo-temporal reconstruction can provide a good approximation of the average development trajectories provided a sufficient number of cells in each cell cycle phase is available for analysis.

Acknowledgement

This thesis would not have been possible without the help of the many people who supported me during the project. I would first like to thank my thesis advisor Ambroise Lambert (PhD) at EPFL for his continuous feedback, helpful discussions and guidance on the biological side of things. I would also like to thank Aster Vanhecke (MSc) and Anna Archetti (MSc) for discussions on programming as well as Kyle Douglass (PhD) and Christian Sieben (PhD) for their support with the microscopy setup.

I am grateful to Prof. Suliana Manley at EPFL for allowing me to explore new paths with unknown outcome and providing feedback at important stages of the thesis. Last but not least I would express my gratitude to Prof. Gerhard Schütz at TU Wien for his support in writing my thesis abroad and providing critical feedback.

Contents

1	Introduction	5
2	Previous work	6
2.1	Cell shape	6
2.2	Single cell identification	6
2.3	Pseudo-temporal classification	7
2.4	Microscopy	10
2.4.1	Diffraction limited	12
2.4.2	Super-resolution	17
3	Temporal reconstruction	21
3.1	Supervised methods	23
3.2	Unsupervised methods	25
3.2.1	Length ordering	25
3.2.2	Diffusion map	26
3.2.3	Wishbone	27
3.3	Cell cycle indicator	27
4	Materials and Methods	28
4.1	Assessment methods	29
4.1.1	Spatial resolution	30
4.1.2	Temporal resolution	30
4.2	Combining trajectories	30
4.2.1	Group based	30
4.2.2	Parametrization	32
4.2.3	Kernel density based	32
4.3	Experimental methods	34
4.3.1	STORM Sample preparation	34
4.3.2	STORM Microscopy	36
4.3.3	SIM Sample preparation	36
4.3.4	SIM Microscopy	37
4.4	Image analysis	38
4.4.1	Localisation rendering	39
4.4.2	Shape	39
4.4.3	Feature importance	43
4.4.4	Temporal reconstruction	43
5	Results	44
5.1	Spatial resolution	44
5.2	Single cell identification	46
5.3	Temporal reconstruction	48

6	Discussion	61
6.1	Spatial resolution	61
6.2	Single cell identification	62
6.3	Pseudo-temporal classification	62
7	Conclusion	63
8	Outlook	64

1 Introduction

Since its invention, the microscope has been used by biologists as a tool to study and describe biological specimen [40]. After observation and classification, research has moved on to propose and study models for cell cycle development [19, 36]. For bacteria, the study of cell shape dynamics proved to be an important tool for establishing and testing a biological cell cycle model [19]. However, for small features like cell invagination before division or protein distribution accurate description is difficult using classical diffraction limited light microscopy.

With the availability of new super-resolution techniques going beyond the classical diffraction limit, even small features can be measured with high precision. However, sample preparation and imaging conditions for some of the highest resolution techniques prevent the capture of time-lapse images of the same cell ruling out direct observation of cell cycle dynamics.

The aim of this thesis is therefore twofold. First, the influence of spatial resolution improvement on the description of the cell cycle shall be assessed. Second, the possibilities to reconstruct cell cycle development from static single cell images of different cells shall be explored. A technique to regain approximate time information (called pseudo-temporal classification) would potentially allow the usage of high spatial resolution methods for the description of cell cycle dynamics.

Interestingly, a number of novel techniques to reconstruct dynamic information of biological developments from static measurements of eukaryotic cells have been developed in recent years [18, 17, 39]. These techniques are based on different types of measurements ranging from single cell RNA sequencing to estimation of the DNA content and produce pseudo-temporal estimations where the measurements are located within the cell cycle. These methods were applied to microscopy data from bacteria and contrasted with traditional methods for approximating the position of a measurement within the cell cycle. Descriptions of the used methods can be found in section 2.3.

Motivation

Super-resolution microscopy technologies like STORM and PALM enable the spatial imaging of cells with unprecedented resolution. However, the usage of these techniques for exploring cell cycle dynamics is limited due to the fact that many of the commonly used sample preparations require the cells to be fixed before imaging, killing them in the process [5]. Since the low spatial resolutions may lead to significant imprecisions in the description of cell cycle dynamics, it would be desirable to estimate the dynamics from a series of high resolution static single cell images.

One approach to this problem is to try to infer the temporal cell cycle state from a combination of static single cell images from a cell population. Under the assumption that measurement noise and the intrinsic cell to cell variability is lower than the cell cycle dependent dynamics, this may enable the reconstruction of the dynamics of a representative cell cycle.

2 Previous work

2.1 Cell shape

The characterization of cell shape is an important requirement for the description of bacteria shape dynamics. Often bacteria shape is described using a combination of geometric forms symmetric to central axis of the cell. For example, a commonly used model of *E. coli* describes the cells as cylinders with hemispherical caps, where the division site is modeled as capped hemispheres [31].

As a rule of thumb, *E. coli* bacteria have dimensions of approximately $2\mu\text{m}$ in length and $1\mu\text{m}$ in diameter [33]. These dimensions are too small for accurate quantitative measurement of cell shape using diffraction-limited light microscopy. The resulting measurement bias by using classical microscopy on small bacteria is described in section 4.1.1.

Building on the model *E. coli*, the dominating form of the *C. crescentus* cells used in this thesis is often described as a cylinder with hemispherical caps [19]. However, in contrast to *E. coli* the *C. crescentus* cells may have significant curvature and grow a long, thin extension at one cell pole called stalk. Based on this model a parametrization for *C. crescentus* is developed in section 4.4.2 and shown in figure 28.

Alternatively, it is also possible to only make general assumptions about cell symmetry and measure the actual shape along the symmetry axis. This path is followed in [41], where the cells are assumed to have a uniform global curvature and cell diameter is measured perpendicular to the symmetry axis (figure 1). The measurement process described in section 4.4 is using similar principles but allows for non-uniform curvature of the symmetry axis.

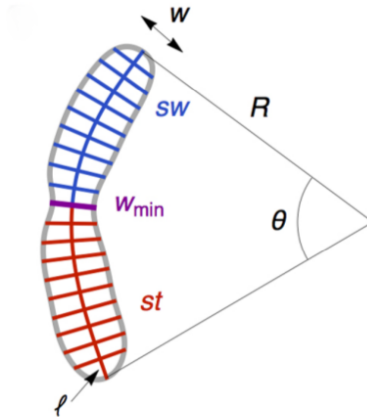


Figure 1: Parametrization of *C. crescentus* along symmetry axis with uniform curvature. Cell width (w) is measured along section perpendicular to the symmetry axis (l). The radius of curvature (R) is shown together with the width at the invagination site (w_{min}). The stalked (st) and swarmer compartments (sw) are shown using color coding. From [41]

2.2 Single cell identification

The observation of cell dynamics without single cell identification can mask important behavior, as population noise may suppress the characteristics of single cell trajectories. With the

availability of new dyes and microscopy techniques, a number of single cell studies have been performed in recent years [31, 41].

Their scopes range from the determination of single cell variability of growth parameters [31] to continuity of shape dynamics between generations of the same cell [41]. While the studies have shown significant variability between single cells, the overall growth behaviour was found to be similar, allowing the construction of a representative population trajectory.

The extent to which the population trajectory of *C. crescentus* represents the single cell behaviour is also discussed in section 5.

2.3 Pseudo-temporal classification

Pseudo-temporal classification tries to infer cell cycle times from measurements of fixed cell populations. Traditional techniques for estimating the number of cells within a cell cycle stage have been known for some time [21]. However, in recent years a number of papers have been published using novel approaches to recover temporal information from fixed eukaryotic cell populations with increased precision. [18, 35, 17].

Two of the methods are mainly based on RNA sequence data, while the other uses nuclear volume, DNA replication and cell cycle markers from microscopy images. They all use various techniques to reduce the high dimensional measurement vector down to a pseudo-temporal scalar.

In contrast, this thesis focuses on cell shape as the main information source for pseudo-temporal classification. This is because unlike eukaryotic cells, bacteria go through a series of highly stereotyped shape changes as they progress through cell cycle. Modifications to the dimensionality reduction techniques described in these papers might therefore be necessary in order to apply them to bacteria cell shape data.

The pseudo-temporal classification techniques can broadly be split into two categories. In the category of supervised methods, new observations are classified based on calibration observations with direct time measurements. This approach requires the availability of a representative calibration dataset and is described in section 3.2.

In contrast unsupervised methods do not require calibration datasets. Instead, they make general assumptions about bacteria development and try to exploit the connections between individual measured bacteria within a dataset. All methods described in the following sections fall into this category.

A general assumption used by all methods is that cell shapes are expected to feature only small differences for each time step. Assuming all phases of the cell cycle are densely sampled, a measure of similarity between cells allows to create a trajectory representing the average cell cycle developments. The methods described below mainly differ in their approach for assessing cell similarity.

Diffusion maps

In diffusion maps the similarity assessment used is the diffusion distance which represents a measure for the probability of transition between the measurement vectors of individual cells. For the propose of this thesis, the measurement vectors contain the parametrized shape of the cells described in section 4.4.2.

The diffusion distance decreases for smaller (euclidean) distances and a larger number of possible transitions between cells. It is therefore robust to noise and perturbations created by outlier cells [7].

A comparison with various similarity measures commonly employed for dimensionality reduction of measurement vectors found diffusion maps to perform best for cell data [18].

Wishbone

Wishbone [35] follows a similar approach as diffusion maps but computes the similarity metric based on the shortest paths instead of transition probabilities between cells. It uses diffusion maps as a preprocessing step to reduce the number of measurement variables in the measurement vector before computing the shortest path between cells.

Another difference to diffusion maps is that instead of calculating pairwise similarities between all cells, wishbone limits the calculations to a fixed number of nearest neighbours, thereby reducing computational complexity and processing time for large datasets. However, it also makes the results more sensitive to outlier cells.

Cycler

Cycler [17] builds on the algorithms developed for Wishbone but is specifically designed to process microscopy measurement data instead of generic feature vectors. Like wishbone it receives a measurement vector as input but modifies details of the algorithm to ensure cells from all cell cycle stages are evenly represented and takes steps to reduce the influence of outlier cells.

In addition to shape measurements of the targeted cells it also takes DNA content and replication information measured using markers as well as local crowding caused by heterogeneous growth environments into account.

Length ordering

Instead of computing similarities based on multiple shape parameters, length ordering focuses on a single parameter for determining shape similarity. With this simplification cell similarity can easily be determined by comparing two scalar numbers.

Ordering cells based on a parameter gradually increasing with time like cell length constitutes a well known approach to pseudo temporal ordering [21]. The technique is described in detail in section 3.2.1.

While overall simple and robust, application in asynchronous populations may lead to biased results [39]. This can be circumvented by applying the time correction described in the next section.

Ergodic rate analysis

Ergodic rate analysis (ERA) uses ergodic assumptions together with the known behaviour of some properties of the cell cycle to construct a pseudo-temporal trajectory from fixed cell measurements. While the use of this technique to quantify cells in discrete cell cycle stages is well established [21], it has been extended in recent years to allow for the description of continuously varying properties such as cell size and DNA content [24].

To connect the dynamics of a single cell to the state of the cell populations, the underlying process must satisfy a criteria known as the ergodic assumption. Following [39], the weak and strong form of these ergodic assumption can be stated for the case of the cell cycle as:

Weak ergodic assumption If the distribution of cells among different states does not change over time, then the proportion observed in any state is proportional to the time each cell spends, on average, in that state.

Strong ergodic assumption If all cells are going through an identical cycle of events, then the proportion of cells in any cycle stage observed in the population at a single time point is the same as the proportion of time spent in cycle stage as a single cell progresses through the cycle.

While the weak ergodic assumption is sufficient for the applicability of ERA the degree of correspondence with the strong ergodic assumption determines how closely the ergodic average represents the behaviour of a single cell [39].

The main idea of ERA is that the number of cells found in a certain state within an asynchronous population is related to the average transit time through that state [24]. If the number of cells grow exponential with a known rate and the behaviour of a continuously varying properties such as cell size is known this can be used to infer the dynamics of other measured quantities. If only an estimation of normalized cell cycle time is desired the exponential rate constant can be left unknown. Since the number of cells doubles at the end of each cell cycle, assuming a uniform correspondence between proportion of observed cells and the cell cycle would lead to an over-representation of cells early in their cell cycle [39]. The cell cycle time is therefore calculated as

$$t = -T \frac{\ln(1 - \frac{p_{\leq t}}{2})}{\ln 2} \quad (1)$$

with t as cell cycle time, T as time to complete one cell cycle and $p_{\leq t}$ as the percentage of cells up to time t [39]. The correspondence between observed percentages and cell cycle time can also be seen in figure 2.

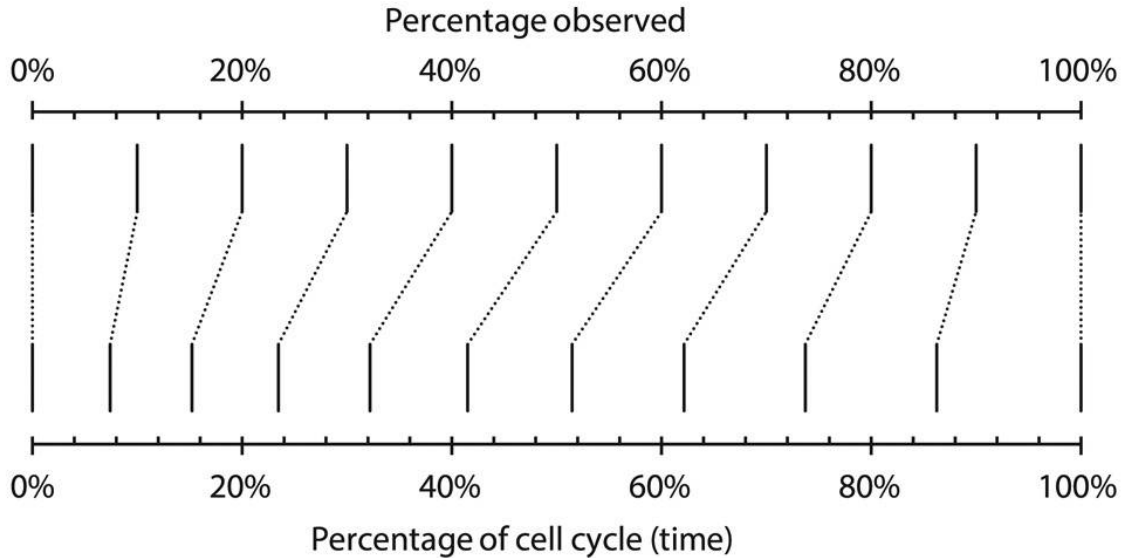


Figure 2: Graphical depiction of the correspondence between the percentage of cells observed up to a given point and the elapsed normalized cell cycle time. From [39].

ERA is also connected to length ordering described in section 3.2.1. Length ordering of an asynchronous population can be regarded as a special case of ERA where the known behaviour is the monotonic increase of cell length. The ability of ERA to correct for the bias of observed cell cycle percentages can be used to improve over simple length ordering in the case of asynchronous population.

2.4 Microscopy

Since over 300 years light microscopy has been used for the observation of biological samples. While initially the resolution of early microscopes was limited by the production quality of the lenses and afterwards by the diffraction limit, new methods have been developed in recent years to overcome this traditional frontier [40].

The following sections briefly introduce the concepts of classical diffraction limited microscopy. Afterwards, an overview of SIM and localization microscopy is given, which are the mainly used microscopy methods for this thesis.

The main limitations of today's classical microscopy is the diffraction limit. First described by Abbe in 1873 [1] it defines a minimum distance between two diffracting points required to reliably separate them. The limit is shown graphically in figure 3. This distance d depends on the light wavelength λ as well as on the refractive index n of the immersion medium and the half opening angle θ between the point and the lens

$$d = \frac{\lambda}{2n \sin \theta} \quad (2)$$

The quantity $n \sin \theta$ is called numerical aperture and is a measure of microscope quality. In modern times good microscopes can reach a value of about 1.4 [10]. For a typically used

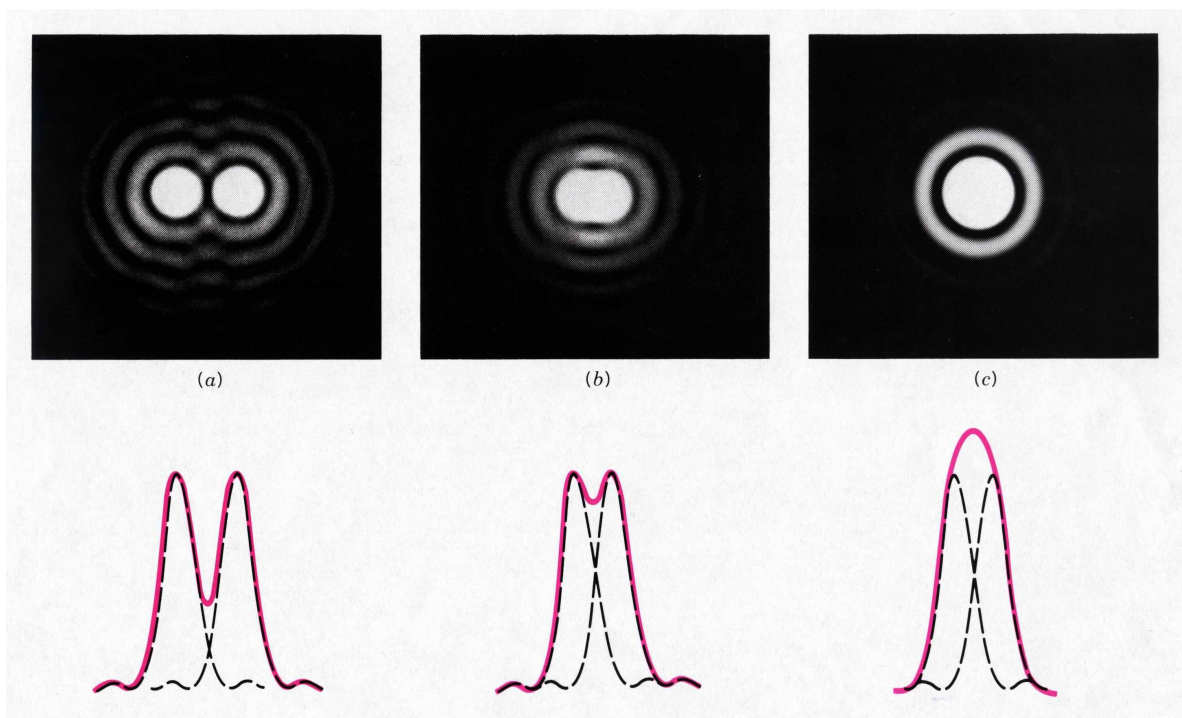


Figure 3: Images of diffraction limited point sources. The top row shows the captured images with the bottom row schematically representing the contributions from the 2 emitters (a) The intensity profiles of the 2 emitters (dotted black lines) can easily distinguished from the superpositioned intensities (pink line) (b) At the resolution limit, the emitters are at the minimum distance required for separation (c) Below the resolution limit, two emitters cannot be separated anymore by the superpositioned intensities. From [32].

wavelength of 500 nm the resolution limit is therefore roughly 180 nm. While the resolution can be improved by lowering the wavelength, the higher energy wave may damage the specimen setting a lower wavelength frontier in practise.

The diffraction limit may also be described by introducing the concept of spatial frequency. Spatial frequency measures how often a sinusoidal spatial pattern repeats itself per unit distance. Therefore, coarse patterns typically have a low spatial frequency and fine patterns a high spatial frequency. For the case of a line grating, spatial frequency is simply the number of lines per unit distance. In this framework the diffraction limit can be understood as the maximum number of lines per unit distance where the line spacing can still be determined. This argument was also used by Abbe in his original derivation of the diffraction limit [1].

For imaging of biological specimen a number of additional challenges need to be considered before choosing a microscopy technique. Due to the similarity between the refractive indices of specimen and surrounding medium a separation between the two regions may be difficult. In addition, it is often desired to image changes in biological composition instead of the raw cell shape. The microscopy methods described below present different trade-offs to solve these challenges.

2.4.1 Diffraction limited

Brightfield

In a brightfield microscope, the specimen is illuminated from below using a condenser. Light is diffracted by the specimen and collected using the objective lens (figure 4). The diffraction reduces the measured intensity depending on the sample density, resulting in a dark image on a bright background, giving the method its name. Compared to other illumination methods the setup is simple and therefore widely used.

However, the method has a number of drawbacks for the imaging of bacteria samples. Since the refractive index of many bacteria is similar to the surrounding medium, image contrast is very low. While image contrast can be improved by staining the bacteria prior to microscopy, this requires additional sample preparation and prevents live cell imaging due to cell fixation for some stains.

In addition out-of-focus material contributes to the measured signal, blurring the focused sample details. Some of these drawbacks are mitigated by the methods described in the following sections at the expense of a more complex microscopy setup.

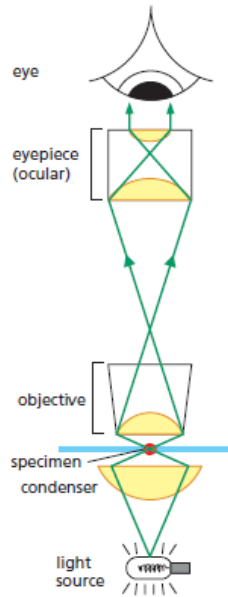


Figure 4: Schematic of a bright field microscope. The light passes through the condenser, is diffracted by the specimen and focused using the objective lens onto the ocular. From [25].

Phase contrast

Phase contrast microscopy enhances contrast for biological samples by separating illumination and sample scattered light using its phase change. Light scattered from the sample typically experiences a phase shift of -90° , whereas the illumination light is unaffected. By shifting of the background light to constructively interfere with the scattered light and dimming the remaining background, sample contrast can be enhanced drastically (figure 5). The method proved to be extremely useful for biology, earning its inventor Frits Zernike the Nobel prize in physics of 1953.

While requiring more components than a simple brightfield microscope, phase contrast microscopy does not require cell staining making it the method of choice to image many otherwise transparent bacteria.

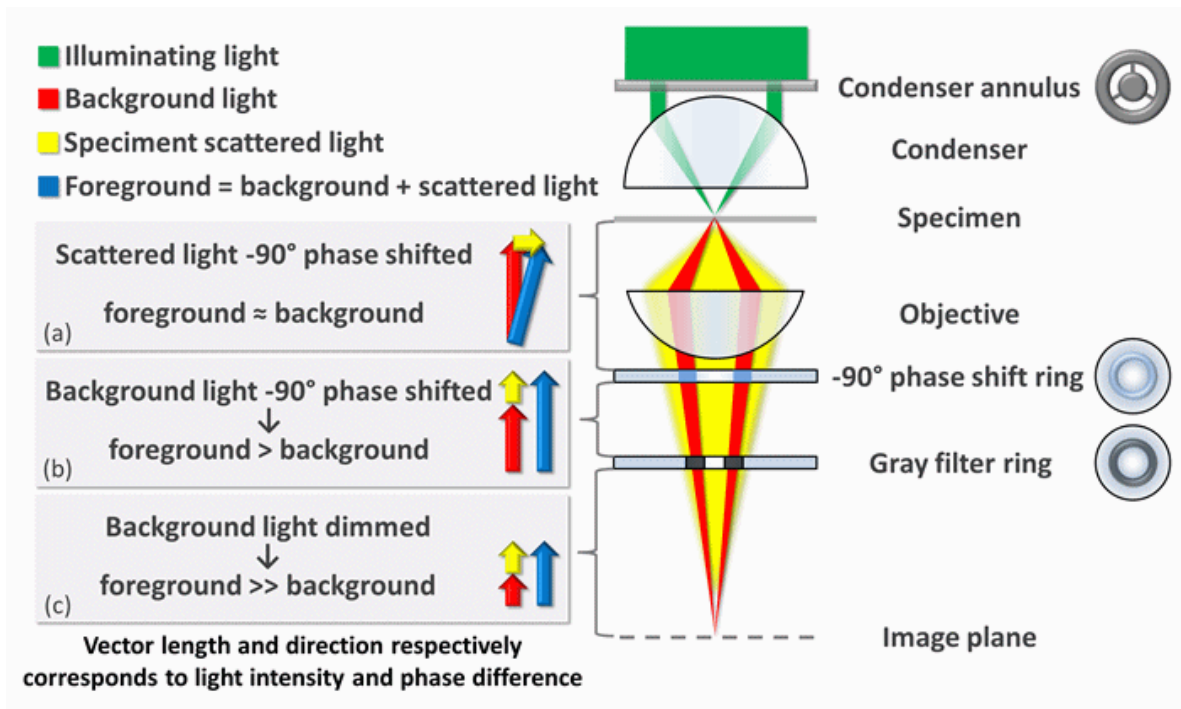


Figure 5: Working principle of a phase contrast microscope. The specimen is illuminated using the condenser and background as well as specimen scattered light is collected by the objective. Afterwards, the phase of the background and scattered light is aligned via a phase shift ring to allow for constructive interference. The final filter ring dims the background light, producing a high contrast image in the image plane. From [13].

Fluorescence

The microscopy methods described above used the scattering of light from the specimen to generate an image. In contrast, fluorescence microscopy obtains information by using fluorescence light emitted from the specimen. The finite size of emission spots generally leads to the same resolution limit as previous methods, however the ability to image specific (fluorescent) properties of the specimen makes it a widely useful method in biology.

A simple setup for fluorescence microscopy is the epifluorescence microscope. Within such a setup illumination and emission light pass the same objective lens, giving the method its name [9].

Light from a high intensity light source (e.g. Xenon arc-discharge lamp or laser) is tuned to the right wavelength by an excitation filter, reflected by a dichroic mirror and focused onto the sample by the objective, emitting fluorescence light. The emitted light passes the objective again, is separated from the reflected excitation light by the dichroic mirror and filtered by the emission filter before it finally reaches the detector. Depending on the specific microscope, the detector might be a camera or a photodetector as in the laser scanning confocal microscope. A schematic of this setup is shown in figure 6.

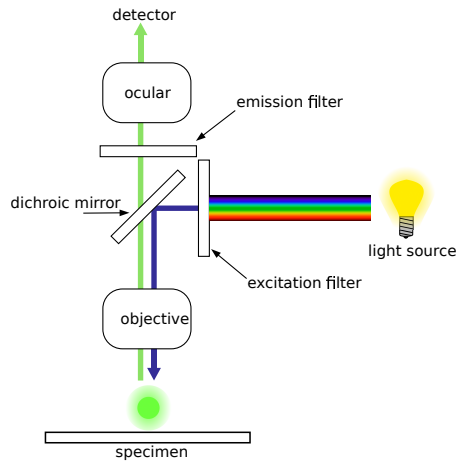


Figure 6: Simplified schema of an epifluorescence microscope. See text further description. From [9].

Since many interesting molecules in biological samples are not naturally fluorescent additional sample preparation is needed before they can be used in fluorescence microscopy. This typically involves staining the sample with a fluorophore. Due to the Stokes' shift, the wavelength of the emission spectrum is usually longer than the excitation spectrum (figure 7) [9]. This allows to separate the two spectra in the fluorescence microscope.

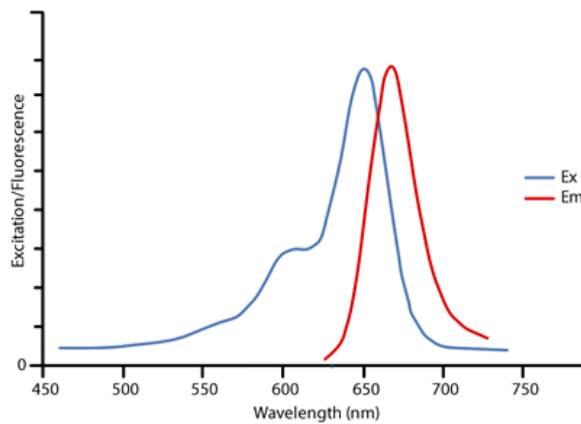


Figure 7: Excitation (Ex) and emission (Em) spectrum for the Alexa Fluor 647. The Emission/Excitation profiles were normalized to peak intensities. In accordance with the Stokes' shift, the emission spectrum is shifted towards higher wavelengths. From [20].

The intensity of fluorescence emission decreases with prolonged exposure to excitation light, an effect known as photobleaching. Photobleaching results from chemical interaction of the fluorophores with molecular oxygen locking them in an excited state and mechanisms of non-radiative energy loss [9]. For time-lapse image series the amount of photobleaching can be a critical factor for image contrast and therefore fluorophore choice (figure 8).

In comparison with previously described methods, fluorescence microscopy requires more sample preparation and a more complex microscopy setup. However, the possibility to fol-

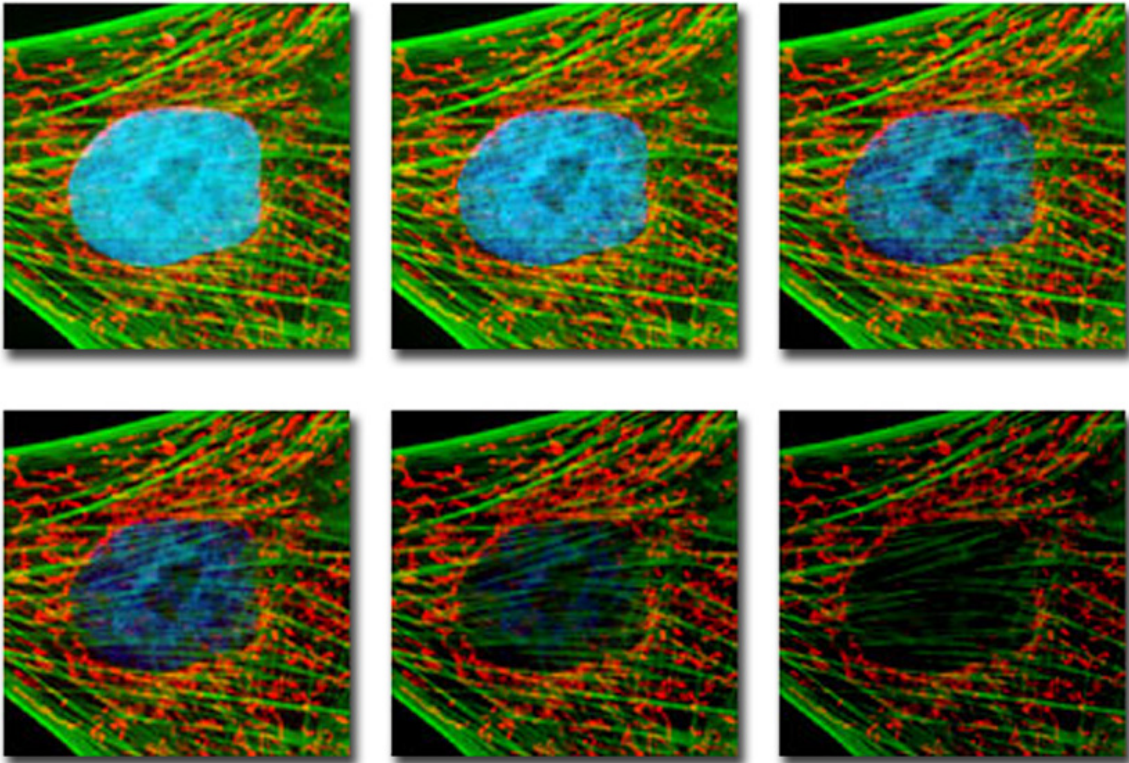


Figure 8: Photobleaching rates of time-lapse fluorescence images of Indian Muntjac deer epidermis fibroblast cells. The nuclei is stained blue (Hoechst 33258),t mitochondria red (MitoTracker Red CMXRos) and the actine cytoskeleton green (phalloidin derivative conjugated to Alexa Fluor 488). An image was taken every two minutes (left to right, top to bottom). The blue fluorophore fades fast and is almost gone after the 4. image while the red and green fluorophores are more resistant to photobleaching. From [9].

low biological processes by tagging involved molecules with fluorophores and record them simultaneously in multicolor images presents a unique advantage, warranting additional effort. Furthermore, some of the methods described below can surpass the diffraction limit, allowing the imaging extremely small biological samples with high resolution.

Laser Scanning Confocal

In confocal microscopy only a small focal volume is illuminated at each time point using a laser beam. A complete image is generated by rastering the sample using a dichroic mirror. The combination of light source and detector pinhole allows to reject out-of-focus light, preventing it from reaching the photodetector (figure 9).

The rejection allows to increase image contrast substantially, especially for thick samples [14]. In addition the methods allows to image optical sections of thick samples, providing depth information.

Its disadvantages are the high cost of purchase and operation compared to traditional microscopes as well as damages to biological samples due to high laser intensities. Being a scanning method, the image acquisition process is also much slower than for traditional microscopes.

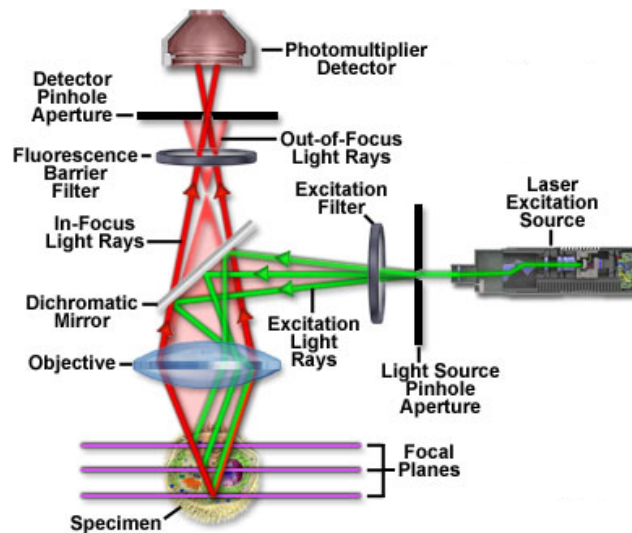


Figure 9: Imaging schematic of a laser scanning confocal microscope. The excitation light passes through a pinhole, is reflected by a dichroic mirror and focused onto the sample by the objective. The fluorescent light generated by the excitation passes through the dichroic mirror, is reduced to in-focus light by the detector pinhole and measured using a detector. The combination of light source and aperture pinhole is responsible for rejecting out-of-focus light. From [14].

2.4.2 Super-resolution

SIM

Structured illumination microscopy achieves resolutions beyond the Abbe limit by exploiting interference between a chosen illumination pattern and the specimen [23]. The interference

results in a mixing between image frequencies, shifting higher frequency information into the observable lower frequency domain. By combining multiple images with shifted illumination patterns, the resulting image contains information from double the observable frequency range, thereby also doubling the resolution. The process required for obtaining these images is outlined in the following paragraphs.

An important technique to describe the combinations of image frequencies is the Fourier analysis. It transforms images from the spatial to the frequency domain, thereby simplifying the convolution of the point spread function (PSF) in spatial domain to a mere multiplication with the optical transfer function (OTF) in the frequency domain (Fig 10).

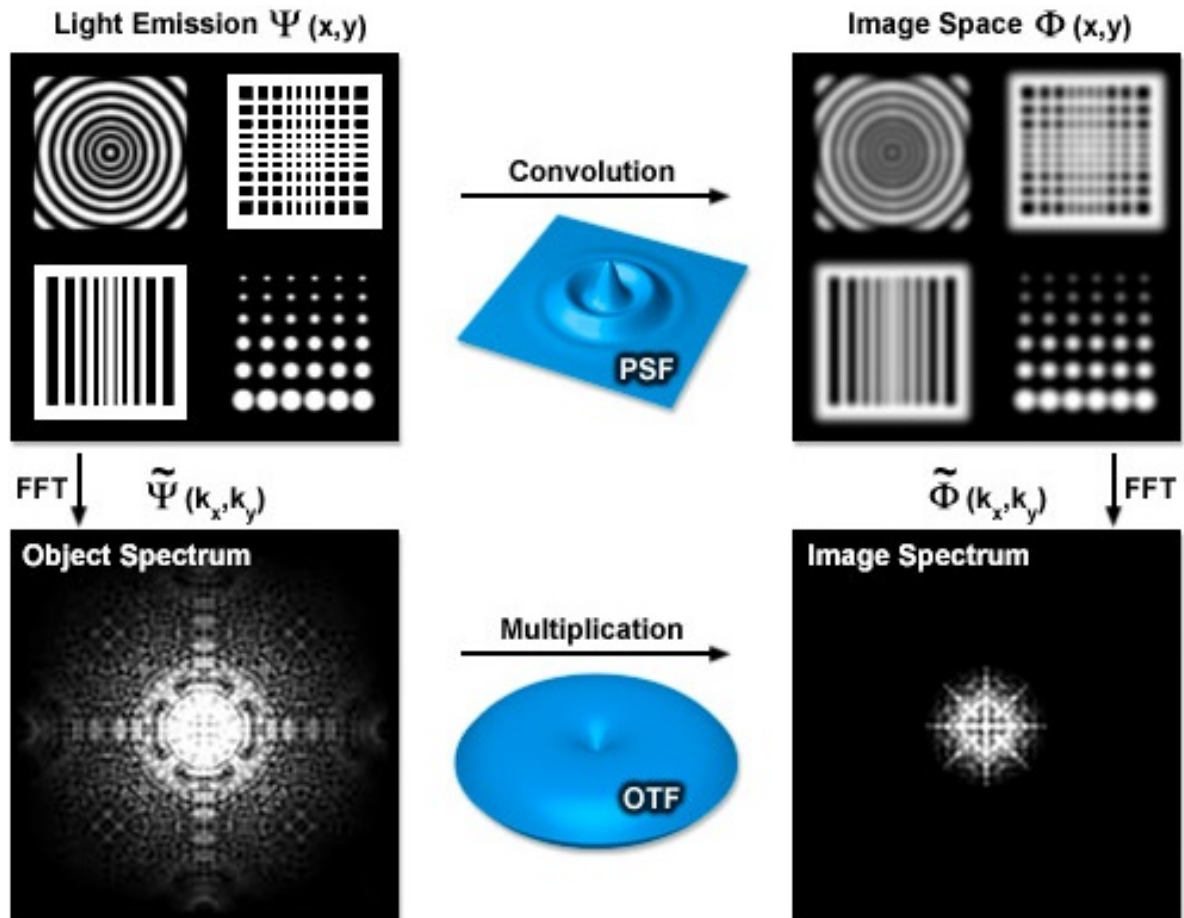


Figure 10: Usage of the Fourier transformation in microscopy. Light emission of the fluorescent molecules in the spatial domain is convoluted with the PSF to produce the observed images. In the frequency domain representation same the observed image is found by multiplying with the OTF. The representation domains domains are connected by the Fourier transform, where the OTF is the Fourier transformed PSF. From [26]

An interference between the specimen and the illumination pattern can be produced by rotating the specimen relative to an illumination grid produced by a grating (Fig 11a). The observable region in frequency space is limited by a circle of radius k_0 as shown in Fig 11b.

With a grating spacing at the Abbe limit, the first order diffraction components k_1 can be used to produce images centered at the boundary of the observable region (Fig 11c). While the directly observable frequency range remains unchanged, frequency mixing carries information beyond k_0 into the observable region (overlapping circles in Fig 11d). Rotating the specimen allows to center the image at successive positions along the k_0 circle, mixing information from all directions of the higher frequency space into the observable region (Fig 11d) [26].

Following [26], this process can be described mathematically as follows. The observed fluorescent signal $E(r)$ at position r is mainly determined by three factors: the local concentration of the fluorophores $D(r)$, the excitation illumination $I(r)$ and the point spread function $PSF(r)$. In the spatial domain $E(r)$ is then calculated as

$$E(r) = [D(r) * I(r)] \otimes PSF(r) \quad (3)$$

where \otimes represents the convolution operation and the PSF models the point-emitters diffraction.

When transferred into a function of frequency k the convolution operation is replaced by a multiplication and is therefore much simpler to compute

$$\tilde{E}(k) = [\tilde{D}(k) * \tilde{I}(k)] * OTF(k) \quad (4)$$

where the tildes represent Fourier transformed functions and the optical transfer function $OTF(k)$ is the Fourier transformed $PSF(r)$. The trick is then to generate a non-uniform illumination pattern $\tilde{I}(k)$ such that the signal $\tilde{E}(k)$ also depends on other frequencies k' . For example, when choosing the sinusoidal pattern with frequency k_0 and phase ϕ

$$I(r) = I_0(1 + \cos(k_0 * r + \phi)) \quad (5)$$

\tilde{E} is a function of $k, k + k_0$ and $k - k_0$. By capturing multiple images as described in the previous paragraph, these contributions can be disentangled, yielding an image resolution beyond the diffraction limit.

An exemplary SIM microscopy setup is shown in Figure 12. Each image is obtained at a grating angle and subsequently transformed into the frequency domain. After all images have been combined, the resulting image is transformed back into the spatial domain using an inverse Fourier transform, yielding the final image.

While SIM microscopy allows to undercut the Abbe limit by a factor of 2, it also has some disadvantages compared to the previously described methods. Multiple microscope image captures are required to obtain a single SIM image, slowing down the image acquisition rate and thereby potentially masking fast transitions [23]. Also, poor estimations for the parameters required by the image reconstruction algorithm may lead to reconstruction artifacts [38]. Despite these drawbacks SIM is widely considered a promising method for microbiology [8].

Localization

Instead of using specialized spatial illumination patterns, localization microscopy modifies the emission behaviour of the sample fluorophores, introducing a time dependency to achieve resolutions below the diffraction limit. This effectively allows to separate the measurements temporally instead of spatially.

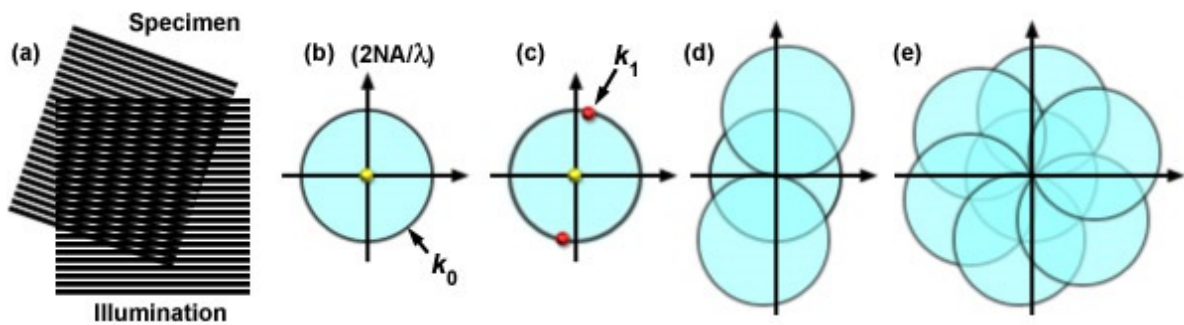


Figure 11: Process of resolution enhancement in SIM. (a) Overlaying the specimen with an grid illumination pattern produces Moiré fringes. (b) The directly observable region is limited to by the Abbe resolution limit to a circle of radius k_0 in the frequency domain. (c) Points at the edge correspond to the first order diffraction components k_1 . (d-e) Relative rotation between specimen and grid illumination allows to mix higher frequencies into the observable region. From [26]

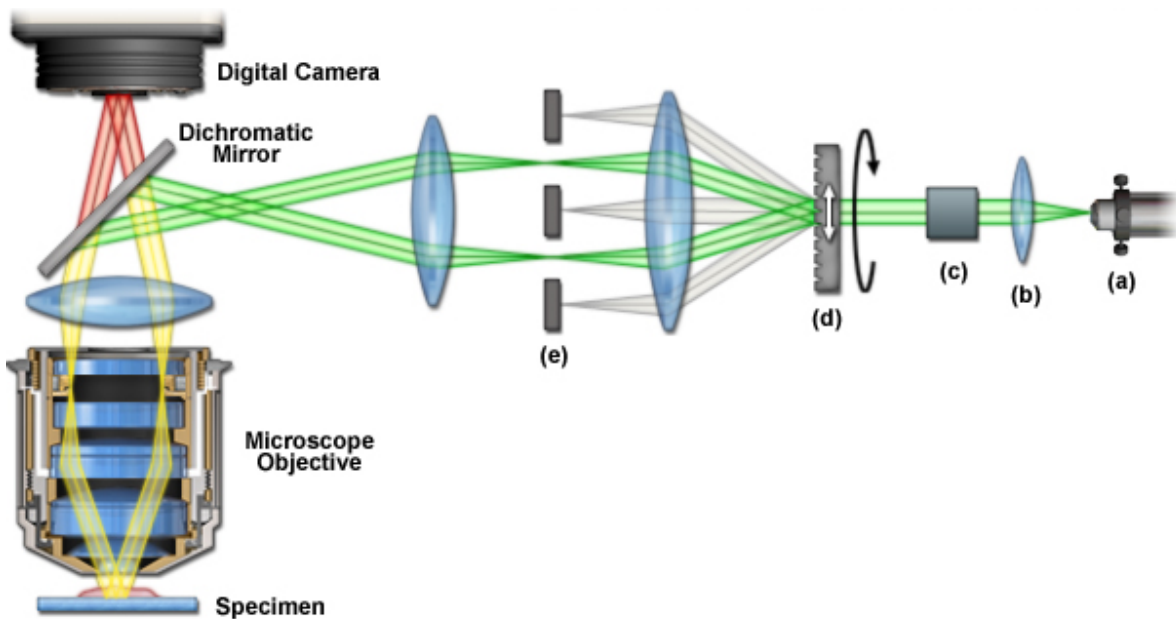


Figure 12: Exemplary setup of a SIM microscope. Coming from from an optical fiber (a), the laser beam is collimated (b) and linearly polarized (c). A phase grid (d) is used to produce a diffraction pattern, of which the first orders are allowed to pass the beam block (e). The resulting beam is focused on the specimen, triggering fluorescence light which passes the dichroic mirror and is captured using a digital camera. Adapted from [26]

The main idea is to limit the emitter density, so that the probability of significant overlap between two emitter point spread functions is extremely low. Under these conditions the position of each emitter can be localized separately without the usual diffraction limits restrictions. To generate a complete picture, different subsets of the fluorophores are activated and localized for each image and then combined to reconstruct the specimens full image. This process is also shown in figure 13.

The achievable localization error σ depends on the number of captured photons N as well as the standard derivation of the Gaussian fit s and can be approximated as [22]

$$\sigma = \frac{s}{\sqrt{N}} \quad (6)$$

An important characteristic for the quality of a dye used for localization imaging is therefore the number of photons emitted before photobleaching occurs.

For achieving a time-dependent emission behaviour of the fluorescent dye, two techniques are widely used. Their main difference roots in the type of used fluorophores.

Stochastic Optical Reconstruction Microscopy (STORM) uses a synthetic dye that is photoactivated by laser illumination. Initially all fluorophores are in the ground state. From there they are stochastically lifted to an excited emitting state by the laser irradiation and fall back to a dark, non-emitting state after a short time. This effectively produces a local blinking behaviour. The emitter density can be controlled by the irradiation intensity, producing the required conditions for localization microscopy.

In contrast, Photoactivated Localization Microscopy (PALM) uses fluorescent proteins for cell staining. Like in STORM, the fluorescent proteins are activated by laser irradiation. However, photo-bleaching confines the activated molecules into the dark ground state after a short amount of time, preventing the emitter density from becoming too high. In general the fluorescent proteins emit less photons than the synthetic dyes used in STORM.

Unlike SIM, localization microscopy has no hard, wavelength depended resolution limit. However, the time required to capture a large number of microscopy images and limited lifetime of the fluorophores limits the resolution in practise to about 30 nm for bacterial samples [11]. In addition, the amount of irradiation damage induced during microscopy makes life cell imaging difficult for many samples.

Localization microscopy is therefore an attractive solution when high resolution images of radiation resistant or fixed cell is desired.

3 Temporal reconstruction

Temporal information is essential for the study of cell cycle dynamics. However, experimental procedures to obtain single cell temporal information using time-lapse microscopy are difficult to implement efficiently in high-throughput experiments [17]. It would therefore be desirable to obtain temporal information without the need for explicit time measurements. Temporal reconstruction methods could provide temporal information as a byproduct of other measurements, potentially enriching existing static measurements with dynamics information.

As a necessary precondition for any temporal reconstruction method, the measured quantities must show a behaviour characteristic for the time at which they are observed. For bacteria the concept of a cell cycle implies such a (periodic) dependency exists. In general

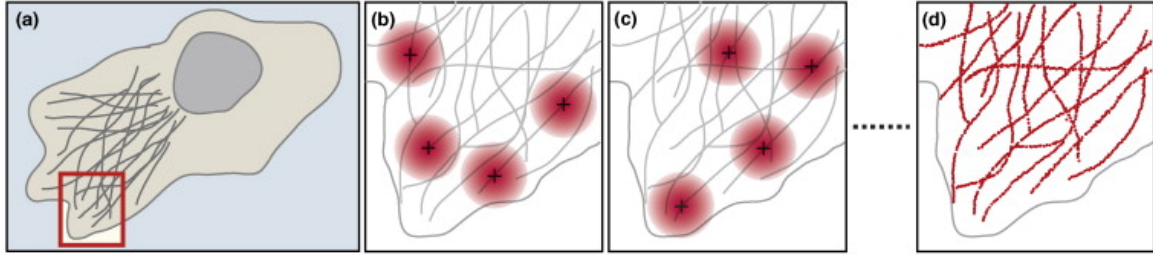


Figure 13: Sketch of the imaging process of STORM, adapted from [4].(a) The selected region of interest is labeled with fluorescent molecules.(b) At each batch a subset of fluorophores is photo-activated and localized (black dots). The density of activation is limited to avoid overlapping of measured signals (red dots).(c) The next batch image measures a different subset of fluorophores.(d) After a sufficient number of localizations were recorded the final image is generated by combining all measured localization data.

the cell cycle duration is different for each bacteria, hence no fixed correlation between cell cycle and global time can be found. Since most biological measurements are expected to have a stronger dependence on the bacteria cell cycle than on global time, all methods described in this thesis primarily target cell cycle time.

The values of the measured quantities vary for different cells at the same cell cycle time. Assuming the individual cell provenance is unknown in the experimental setup, the change over time should be greater than fluctuations within the population at a fixed cell cycle time point. Otherwise these fluctuations would mask the time dependence of the measured quantity and make temporal reconstruction impractical.

A biological quantity which is readily available using microscopy is cell shape. For *C. crescentus* the cell shape undergoes drastic changes, doubling its length and separating into two compartments over the course of a cell cycle [36]. It is therefore a promising quantity to use for temporal reconstruction. Nearly all methods described in the following sections are primarily based on cell contour measurements.

The distribution of the measured quantities may provide additional information useful for temporal reconstruction. Combined with unbiased sampling of the individual cells and basic assumptions about monotonicity or a measure for shape proximity it forms the basis of most of the temporal reconstruction techniques described in the following sections.

Another approach would be to use specialized markers with known changes over the course of the cell cycle. While this approach can be used determine cell cycle stages it requires specialized measurements in addition to actual target of the study. Furthermore, the time dependence of the marker has to be known beforehand and may change between cell types.

Generally the temporal reconstruction methods can be divided into two categories depending on the information they receive for reconstruction. Unsupervised methods are only provided with a dataset containing no direct time measurement information. In contrast supervised methods receive an additional calibration dataset with measured times and perform temporal reconstruction based on this information. The benefits and drawbacks of both approaches are described in the following sections.

3.1 Supervised methods

Supervised methods for temporal reconstruction build a temporal classification function from samples with measured time. This function is then used to infer time for samples for which no direct time measurements are available. For *C. crescentus* such time measurements can be obtained by performing time-lapse SIM experiments.

While these methods may enable temporal reconstruction for complicated time dependencies, they require the assumption that the behaviour of the samples used for training is similar enough to the samples used for classification. When comparing different bacterial strains this condition may be violated requiring a new calibration for each strain. The ability to find a sufficiently general classification function is therefore essential for the applicability of supervised methods.

Average calibration trajectory

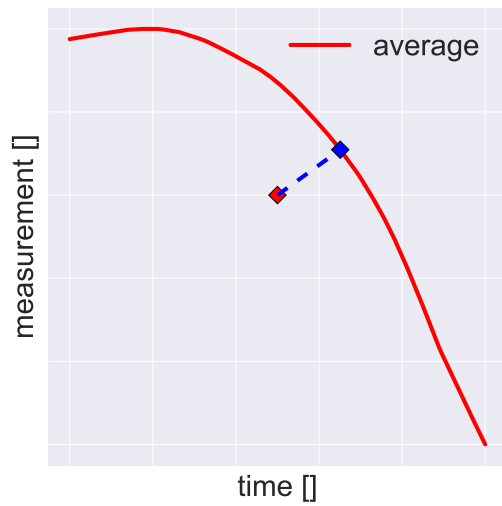
A rather straight forward approach for obtaining a classification function from measured times is to fit a smooth average to the time dependent measurements. The resulting calibration trajectory can then be used to obtain times for new measurements by finding the closet point on the calibration curve to the new measurement. Closeness can be determined by any suitable distance metric, e.g. the euclidean distance. Normalization of each dimension before distance calculations ensures that differences in measurement ranges do not affect the result. This process is shown graphically in figure 14a. By assigning weights to the distances in each dimension differences in measurement uncertainties for each variable can be taken into account.

Nearest Neighbours Regression

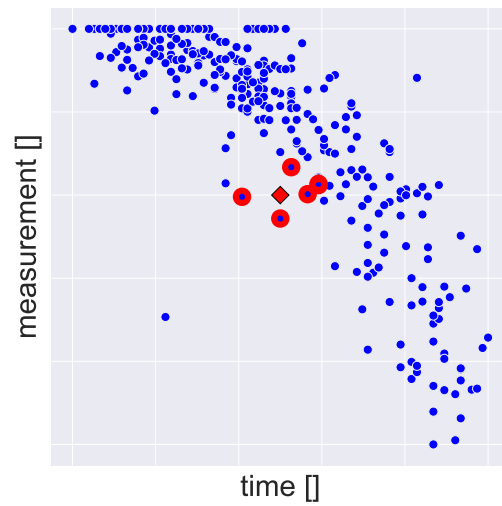
The method used in the previous section required computing an average over the whole data set in advance, reducing the measurement scatter to a single trajectory. Instead of performing this information-reducing step, it is possible to retain the information of the measurement distribution by classifying new points based on local proximity to training points. This method is called nearest neighbour regression and is preformed by averaging over the measured times of the k closet points. The average can either be uniform or weighted to put emphasis on the local structure near the classified point. Like for the calibration trajectory, differences in measurement uncertainties can be easily taken into account by adjusting the weights of distances in different measurement dimensions. The basic principle of this method is shown in figure 14b.

Random forest classification

Instead of obtaining a classification trajectory by computing a time average, machine learning algorithms like random forest may be used to learn a more complex model from the measured times. Random forest fits a number of decision trees to the data, each using a different subset of measurements. While decision trees can capture complex relation between measurement variables, they are also known to overly adjust to the specifics of their training data set, a process called overfitting. Random forest aims to reduce this overfitting by averaging over



(a) Average trajectory



(b) Nearest neighbour

Figure 14: Schema of the classification process using different methods (a) The calibration trajectory (red line) is constructed by a smoothing function (lowess smoothing) over all measurements with known time. A measurement with unknown time (red spade) is assigned the time of the closest point on the calibration trajectory (blue spade). (b) A point with unknown time is classified by averaging over the times of the k nearest points (red circles) from all measurements with known time (blue scatter)

the predicted values from all trees in the forest, therefore reducing the influence the bias of an individual tree may have on the end result.

3.2 Unsupervised methods

Unsupervised methods have to work with much less information than their supervised counterparts. They are not provided with a training dataset with known times, instead they only receive the measurements with unknown times for classification. In order to provide classifications with this information they have to make assumptions about fundamental behaviour of the measurement trajectories.

These assumptions are usually based on known biological behavior (e.g. cells elongate over time) or morphological changes (e.g. similar cell shapes are typically close in time). Different techniques of utilizing these assumptions are described in the following sections.

While these assumptions are qualitative in nature, they are more likely to hold between different cells and stains than statements about quantitative behaviour. It therefore interesting to explore if their pseudo-temporal classifications are precise enough to allow the construction of meaningful population trajectory for different strains.

3.2.1 Length ordering

Most bacteria elongate over the course of their cell cycle [21]. A quantity naturally coming to mind for temporal reconstruction is therefore cell length. Since the elongation over the course of a cell cycle is rather large, it is easy to measure and quantify. For *C. crescentus* this length should exclude the cell stalk as it shows a growth behaviour different from the rest of the cell [36]. Since cells staying within certain size limits have a larger change of survival, their size tends to be similar, leading to a phenomenon known as cell size homeostasis.

As cells are known to grow monotonically over the course of their cell cycle, small lengths tend to correspond to cells early in their cell cycle. However, since the start and end cell lengths may fluctuate between cells, the precision of this approach depends on the fluctuations within the cell population. Fortunately cell size homeostasis implies reasonable limits for these fluctuations, making the approach feasible.

A temporal reconstruction based on cell ordering requires assumptions in addition to the measured cell length in order to yield predictions. One could assume the length time correspondence function is known before hand. Alternatively one could assume the sampling bias is negligibly and the length growth function is monotonic. Both approaches are described in more detail in the following paragraphs.

Since the length time correspondence function maybe unknown a priori, a simple correspondence function to assume could be linear interpolation between the smallest and largest observed values. However, the growth of many cells is not linear and the approach suffers from outliers at the start and end of the cell cycle. Even if the functional form of the growth is known a priori its assumption may bias the resulting trajectories, making it harder to detect variations between populations.

If the sampling bias is negligible it is possible to proceed without assuming a concrete functional form. It is sufficient to know function is monotonically ascending. The cells can then be sorted by length, assigning the smallest cell to the beginning and the largest to the end of the cell cycle (figure 15). This effectively redistributes the cells to an evenly spaced grid

where the positions depend only on the ordering and not the absolute difference between cells. Because of negligible sampling bias, regions with slower growth will have more cells, decreasing the predicted growth rate in this region. Cells exhibiting an atypical growth behaviour only affect the predicted function locally. Their influence is particularly strong in the beginning and the end of the cell cycle as unusually small or large cells are placed there.

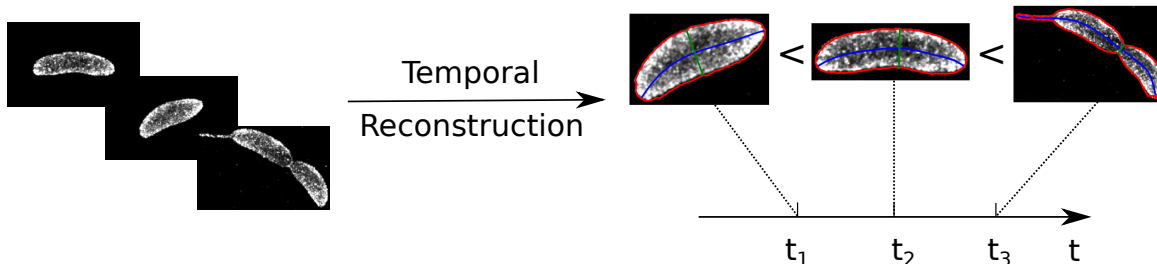


Figure 15: Schematic of the length ordering process. From a stack of images (left) shape measurements are extracted (right). The shapes are then ordered by their measured length (blue line) and mapped to the time axis.

3.2.2 Diffusion map

Diffusion maps are based on the observation that the measurements are not distributed randomly, instead they form a path in higher dimensional measurement space. By simulating a diffusion process between measurements, a measure for their local similarity can be calculated. With assumptions where the start and end of the cell cycle process is located, following this path allows to construct a pseudo-temporal average based on measurement similarity. Cell shape parameters are good candidates for a similarity comparison as they are expected to change slowly over the course of the cell cycle.

The process of generating a diffusion map can be described as follows. First, each measurement is assigned a (Gaussian) kernel function fixed width in measurement space. Then, the interference between all measurement kernels pairs is calculated as a measure for transition probability between cells. An eigenvalue decomposition on the resulting matrix of pairwise transition probabilities is then performed. The first eigenvector corresponding to the first eigenvalue is then a measure of the dominating component of the path, which corresponds to the cell cycle trajectory. To obtain cell cycle time, the components of the first eigenvector are normalized for a range between 0 and 1.

The choice of kernel width has a strong influence on the resulting diffusion distance. If the width is too small, the resulting trajectory may contain gaps. A too large width may connect unlike cells so that the differences between different parts of the cell cycle may be blurred. The choice of a suitable kernel width is therefore essential for good results. Following [18], the kernel width was estimated by the average intrinsic dimensionality of the measurements in this thesis.

Diffusion maps require a densely sampled data set to produce a continuous trajectory. In addition the trunk of the path generated by the cell cycle must be distinguishable from the measurement noise. Under these conditions, the method represents an attractive option to estimate a cell cycle trajectory from high dimensional data.

3.2.3 Wishbone

In contrast to generic dimensionality reduction techniques like diffusion maps Wishbone is a method developed specifically for pseudo-temporal reconstruction of single cell data [35]. It uses a k-nearest-neighbours graph and a Gaussian distance metric to measure local proximity. The shortest path between two cells is then taken to be their pseudo-temporal distance. Initially all distances are computed relative to a predefined starting cell. To reduce the influence of the starting cells, a random subset of all cells is subsequently chosen as waypoints and a new pseudo-temporal distance is computed as the weighted average of their shortest paths to a given cell. This process is repeated until the obtained distances converge. A graphical description of this procedure can be found in 16.

The authors claim the method is robust against noise and the choice of starting cell [35]. A version of wishbone called cyclus [17] has been developed, modifying the algorithm for microscopy data by changing the waypoints selection mechanism to ensure cells from all cell cycle stages are represented.

While wishbone was initially intended for mammalian cells its design should allow it to function for bacteria cells as well. It also has the ability to identify bifurcations in the bacteria trajectories, a feature which is not required for the scope of this thesis.

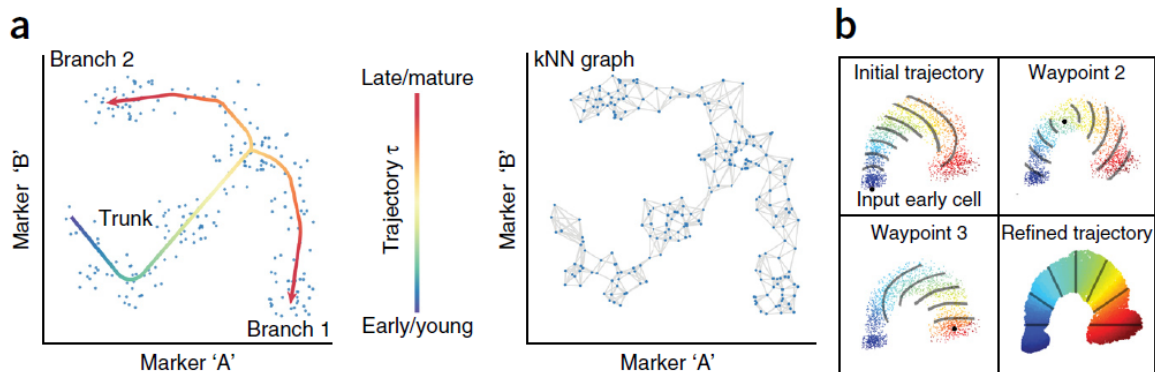


Figure 16: Wishbone pseudo-temporal ordering of cells. (a) The wishbone path preferentially follows the lower dimensional path embedded in the higher dimensional space. It also allows to detect a branching point in the trajectory (left image). Wishbone determines the distances between two cells by computing the shortest path connecting them in the k-nearest-neighbours network (right image). (b) The initial distances computed from a defined starting cell (top left) are refined by the weighted averages to randomly chosen waypoint cells (top right, bottom left) to yield the final pseudo-temporal ordering (bottom left). The colors indicate the pseudo-temporal ordering with the black contour lines corresponding to cells of similar time. From [35]

3.3 Cell cycle indicator

The progression of bacterial cell cycle is driven by a complex genetic machinery. For *C. crescentus*, the cell cycle can be divided into 3 major stages: the G_1 , S and G_2 phase. During the G_1 phase the cell features a flagellum and pili, allowing it to stay motile and explore the

environment. At the transition to the S phase the cell drops the flagellum and pili, instead growing a stalk and becoming sessile. The stalk is a cylindrical structure with a strong adhesive at its tip, allowing the cell permanently attach itself to a surface. During this phase the DNA is replicated and the cell starts to divide. The chromosomes for the daughter cells are separated and the cell grows a new flagellum at the end opposite to the stalk in the G_2 phase. This process is also depicted graphically in figure 17.

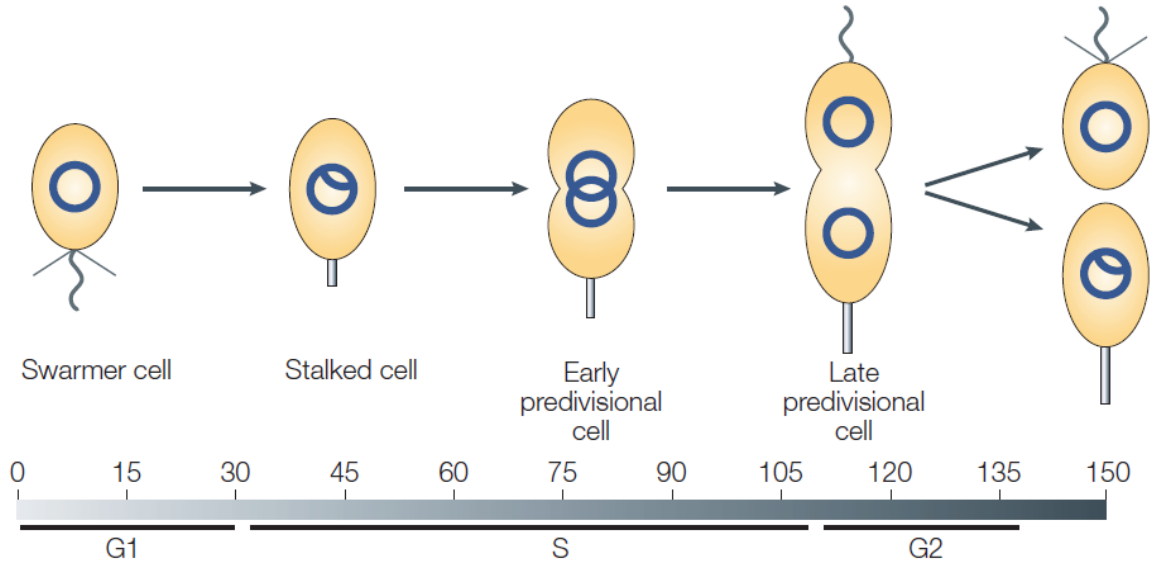


Figure 17: The cell cycle phases of *C. crescentus* with schematic indication of their form. Circles represents the quiescent, ' θ ' the replicating chromosomes. The time line representing the cell cycle time is in minutes. From [36].

It follows that an attractive cell cycle indicator to temporally resolve the development during the S phase could be the total DNA content of the cell. Since this indicator relies on features common in most bacteria cell cycles it is applicable to a wide range of bacteria. However, it only describes a fraction of the cell cycle and must be combined with other indicators to yield predictions for the full cell cycle. An implementation of this technique is described in section 4.4.2.

Alternatively cell cycle indicators specific for *C. crescentus* could be used. Good candidates would be the three so called master regulators CtrA, DnaA and GcrA which are known to drive the progression of the cell cycle [16]. As seen in figure 18 these proteins are periodic within the cell cycle. Since their functions are not invertible for most of the cell cycle additional context would be required to use these proteins as indicators for the pseudo-temporal reconstruction of the full cell cycle.

4 Materials and Methods

C. crescentus is often used as a model organism to study bacteria cell division [16]. The bacteria is easy to synchronize, has a stalked and swarmer cell type with different functions and divides asymmetrically. In addition, it only has a single DNA replication point and a relatively short doubling time.

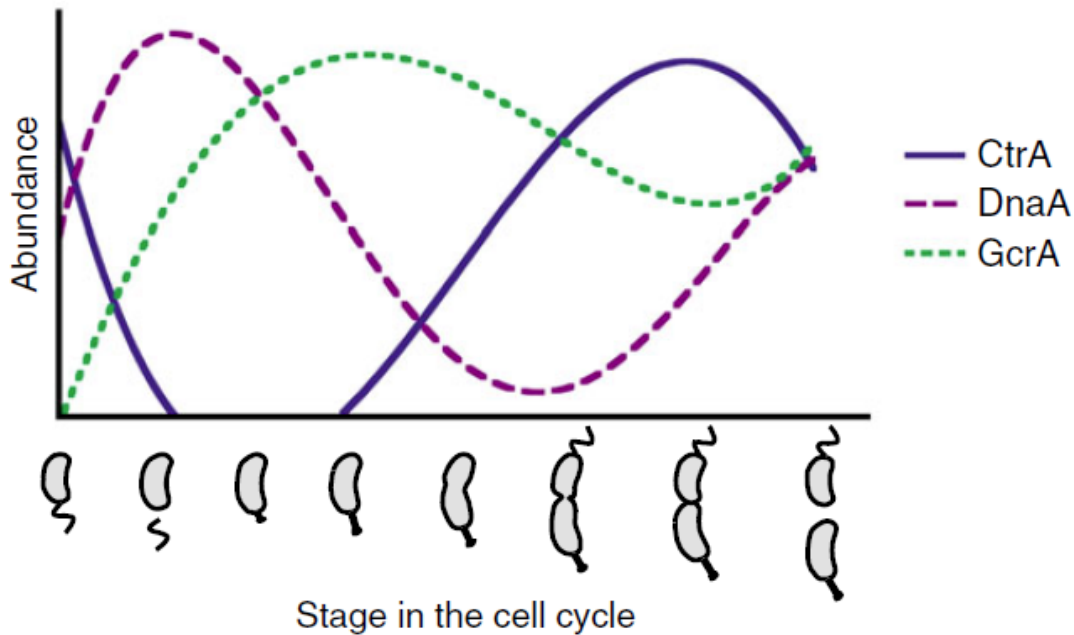


Figure 18: Schematic abundance of 3 proteins driving the progression of the *C. crescentus* cell cycle as function of cell cycle stages. These proteins are also known as master regulators. From [16].

The distinction between stalked and swarmer cells allows for a separation of concerns. Swarmer cells are motile and can search the environment for new nutrition sources. In contrast, stalked cells can use the adhesive holdfast to permanently attach themselves to a surface. Only they can initiate chromosome replication and therefore progress towards cell division. Swarmer cells are able to differentiate into stalked cells after a period of motility, completing the cell cycle. This distinction allows the cells to split the effort searching for new and exploiting known nutrition sources, increasing the probability of survival.

All these properties make *C. crescentus* an attractive organism for studying bacterial dynamics and searching for patterns in this biologically complex process.

4.1 Assessment methods

In order to compare the different microscopy and temporal reconstruction techniques an expressive comparison metric is required. For spatial measurements, the main quantity of interest is the amount of captured details and introduced bias by lower resolution measurements.

For temporal resolution, the metric is more difficult to define. Since only a single measurement for each cell is generally available, the method can only be expected to reconstruct population averages. Because of biological variability, cells measured at the same time may lead to different measured values. A simple difference between predicted and observed time on a single cell level therefore provide an incomplete picture. A comparison between calculated population averages is therefore the main method of comparison for assessing the temporal resolution of the pseudo-temporal reconstruction methods.

4.1.1 Spatial resolution

A simple approach to assess the influence of spatial resolution on shape measurement is to measure cells simultaneously with multiple microscopy techniques. Differences in the measured shapes can then be attributed directly to resolution differences. If the experimental procedures do not permit the usage of all microscopy methods simultaneously, the usage of different cells within the same developmental stadium may allow for a similar comparison.

4.1.2 Temporal resolution

To accurately describe cell dynamics, the main quantity of interest is the difference between the temporal averages of the measured quantities. Since the goal is to observe the overall population dynamics, this measurement provides more meaningful information than the simple sum of time differences between all single cell measurements.

4.2 Combining trajectories

To create temporal averages of population trajectories, a method for combining single cell measurements has to be chosen. Such a method is also required where single cell trajectories with direct temporal measurements are available as cell to cell variability may otherwise mask population behaviour.

The methods can be roughly divided into group based and kernel density based methods. Group based methods compute the population trajectory by applying a function to a subset of all measured cells simultaneously. In contrast, kernel density based method assign a function to each cell individually and compute the population average by summation over the single cell functions. Both approaches have their advantages and drawbacks, described in the following sections.

The choice of method may also depend on the dimensionality of the measurement variable. For example, cell contour is inherently a 2D measurement requiring the interaction of two dimensions for a meaningful cell representation. If the two dimensions were averaged separately and then combined, the result would not represent a meaningful average.

4.2.1 Group based

Group based methods typically use a sliding window to split the dataset into multiple subsets. A function is then applied to each subset, generating the averaged value. The sliding window size determines the stability and lagging of the result. A bigger sliding window typically leads to more stable average at the expense of increased lag to local changes of the dataset.

Group based methods therefore require a dense measurement for a stable and responsive average. In addition the measurement should not contain spikes or discontinuities, as these may be masked in the averaging process.

Group based methods are widely used as they allow easy interpretation and computation of confidence interval. However, their result may strongly depend on the sliding window size, systematically shifting inflection points for larger window sizes. They are therefore attractive when the measurements are known to vary slowly and measurement density allows for a sufficiently small window size.

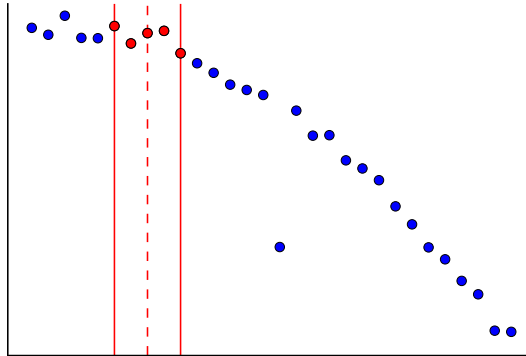


Figure 19: Illustration of the sliding window concept. A fixed number of points is selected (red points) and the computed average is assigned to the center of the window (red dashed line). The window boundaries (red lines) are then shifted to compute the next average.

Mean

The simplest function for average computation within the sliding window is the arithmetic mean. It is parameter free and typically works well when the spread of the measured values is small. However, a small number of outliers may be sufficient to bias the result, leading to a discontinuities in the resulting average. If outliers are expected in the measurement using a median may lead to more stable results (figure 20)

Median

Like the mean, the median can be used as a simple, parameter free average function. In comparison to the arithmetic mean it is more robust to outliers but may depend more on the sampled values. This may cause fluctuations leading to a non smooth average.

Lowess

The methods described before used an average computation on the sliding window data. All measurements within the sliding window are treated as a list, discarding information that may be contained in the local structure of the measurements. However, the measurement scatter may also be treated as a local regression problem, assigning each window a value from a local fit.

This approach is taken by the lowess (locally weighted scatterplot smoothing) method. For each window a (linear) fit is performed, where the individual points are weighted by a function depending on their distance to the currently estimated positions (typically a tricube function). This allows to approximate even a non-linear function with a linear fit. As for the previously described methods, a larger window size leads to a more stable curve with less responsiveness to local changes.

In comparison to the previously described methods lowess requires significantly more computing power and densely sampled measurements. Due to its fit based approach it is able

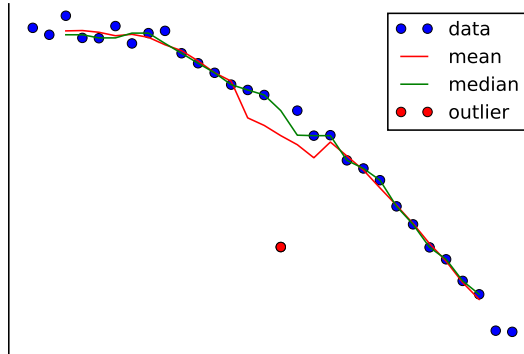


Figure 20: Comparison of the sliding window average with 5 points generated by mean and median computations. The mean is strongly affected by the single outlier point (red), while the median is only slightly deviated. Since at least 5 points are required for the average computations, the averages do not cover the full data range.

to better exploit the information contained in the local structure of the measurements. The results may therefore be closer to what a human would consider a suitable smooth average. If the dataset is densely sampled and sufficient computational power is available, lowess can provide smooth trajectories representative for population behavior.

4.2.2 Parametrization

For two dimensional measurements, the previously described methods may yield non-representative averages due to interactions between the measurements dimensions. This is shown in figure 21, where differences in cell length leads to distortions in the right cell poles contour. One approach to solve this problem is to define a functional parametrization of the expected shape and average the shape parameters instead of the raw shape.

The one dimensional shape parameters can then be averaged using any of the previously described methods. The procedure to fit the parametrization to the measured shape is described in section 4.4.2 with the parametrization for *C. crescentus* being shown in figure 28.

Shape parametrization is a trade-off between preservation of characteristic shape features and bias introduced by the predefined functional form. For the purpose of this thesis the need to preserve shape features dominated, prompting the use of parametrization for the description of cell shape dynamics.

4.2.3 Kernel density based

Kernel density based methods estimate the behaviour of a function by assigning each measurement a kernel function. The kernel is centered at the measured value and the overall function is computed by summation over all kernel values at a particular point (figure 22). In contrast to group based methods, this results in the estimation of a kernel density and does not require the definition of a window size.

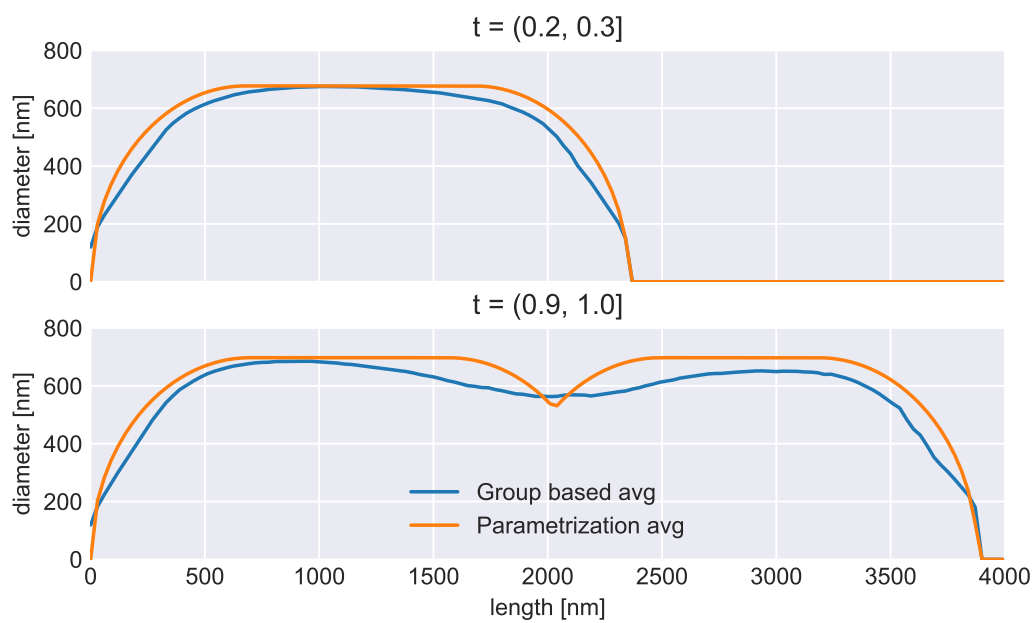


Figure 21: Comparison of the shape combination results using group based mean and parametrization average of 2 cell cycle time bins. The group based method follows the measured shape more closely but is affected by noise at the right pole resulting from different cell lengths. In addition the invagination site is partially "smoothed over" in the shape for the second time bin. The parametrization average is less noisy at the expense of a simplified shape description.

Instead, the choice of kernel type and width determines the smoothness and dependence on local variability of the resulting function. A popular choice is the Gaussian kernel and the width can be chosen to represent the measurement uncertainties introduced by factors like spatial resolution.

In comparison to group based methods the choice of kernel parameters is straight forward while the interpretation of the results can be more difficult. The density cannot be directly converted to an average trajectory, instead more complex approaches like following the peak density of a two dimensional kernel density need to be applied. In comparison to average based group methods the kernel density is also more expensive in terms of computational power.

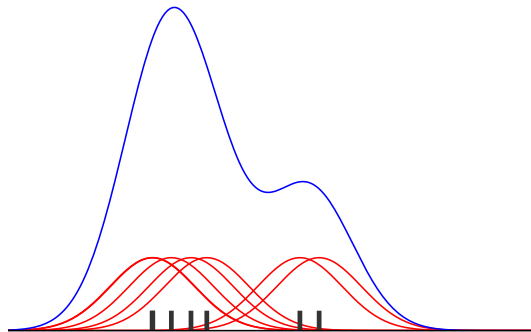


Figure 22: Illustration of kernel density estimation. A (Gaussian) kernel function (red) is placed at the position of each measured value (small black bars). By summation over all kernels a density estimate for the full data range is obtained (black line).

4.3 Experimental methods

In this thesis, mainly two microscopy methods were used to image the shape of *C. crescentus*. By using SIM time-lapse microscopy shape dynamics could be observed directly. STORM microscopy provided a higher resolution images, but sample preparation required cell fixation, necessitating the use of pseudo-temporal reconstruction methods for shape dynamics.

Both methods required staining of the cells with fluorescent dyes, a process described in the following section.

4.3.1 STORM Sample preparation

Cell growth

Under sterile conditions, three CB15N colonies were taken from a petri dish, suspended in 50 ml of PYE (Peptone Yeast Extract) and grown overnight at 28°C in an incubator. If the ODE density was greater than 0.04, PYE was added and the bacteria solution was diluted to an ODE density of 0.03 and grown for an additional hour.

The resulting solution was centrifuged for 5 min at max. speed. Afterwards the supernatant was removed leaving the bacteria pellet on the bottom of the flask and resuspended in 900 μ l of M2 salt previously cooled on ice.

Cell synchronization

All of the synchronization procedure was performed under sterile conditions. The solution was again centrifuged for 3 min at max. speed, resuspended in 900 μ l of M2G and then split into separate tubes of 300 μ l each. Afterwards, 600 μ l of cooled M2 and 900 μ l of cooled percoll were added to each tube.

The tubes were centrifuged for 20 min at max. speed at 4°C. The density gradient of the solution lead to two bands of bacteria in each tube, with the upper band containing the stalked and the lower the swarmer cells. The upper band was pipetted out and discarded. The lower band cells were pipetted out using a 200 μ l pipette and combined into a single cone.

200 μ l of cells were taken out for immediate fixation. The rest of the cells together with 1 ml of PYE was put into a glass tube and grown in the incubator. At the chosen time points, 200 μ l of the bacterial solution were extracted out of the glass tube to proceed with the steps described in the next section.

Fixation & permeabilization

1 ml of bacteria solution was centrifuged for 3min. Under sterile conditions, the bacteria pellet was resuspended in a PBS (Phosphate-buffered saline) and 2.5% PFA (Paraformaldehyde, fixation) and 0.1% Triton (permeabilization) while wearing nitrile gloves. After 30 min the solution was centrifuged again and the pellet resuspended in 250 μ l PBS.

Staining

For staining 0.5 μ l picogreen (DNA dye) and 25 μ l WGA (Wheat germ agglutinin, membrane dye) at room temperature were added to the solution. After resting for 30 min the bacteria were washed 5 times by centrifugation and subsequent resuspension in 250 μ l PBS.

Buffer

To increase the photon yield per fluorophore, the bacteria solution was suspended in a specialized buffer. The buffer reduced the amount of photo bleaching allowing to increase the amount photons available for localization and therefore improving localization precision in accordance with equation 6.

The buffer was prepared following [27]. It consisted of 10 mM MAE(Mercaptoethylamine), 50 mM BME (β -mercaptoethanol), 2 mM COT (Cyclooctatetraene) in 25% PBS, 2.5 mM PCA (Protocatechuic acid), 50 nM PCD (Protocatechuic dioxygenase) combined with 75% glucose (w/v) solution.

Cover slip

An agar solution of 2% was prepared and heated up in a microwave until the agar was fully dissolved. The solution was then filled into a disk-shaped enclosure on top of a cover slip and

sealed using a second glass pad, avoiding the production of bubbles. After 3 min the top pad was removed using a scalpel.

At the same time the bacteria solution was centrifuged for 5 min at max. speed and resuspended in 50 μ l PBS. The freshly prepared COT buffer was added and the suspension remixed.

1 μ l of the solution was pipetted onto the middle of the agar pad, making sure to avoid touching the agar with the pipette tip. After resealing the cover slip with a top glass pad, the sample was ready for STORM microscopy.

4.3.2 STORM Microscopy

STORM microscopy was performed on a novel microscope, optimized for flat-field illumination and a large field of view [11]. This allowed to capture a large number of bacteria per field of view, speeding up the image acquisition process.

The design is based on a epi-illumination microscope and uses a Köhler integrator like system to achieve flat-field illumination (figure 23).

The setup was equipped with a 642 nm laser (2RU-VFL-P-2000-642-B1R, MPB Communications) for sample illumination and a 405 nm (OBIS, Coherent) controlling the return rate of the fluorophores. Excitation light was reflected by a dichroic mirror and focused by an objective (CFI60 PlanApo Lambda x60/NA 1.4, Nikon) onto the sample. Emission light was collected using the same objective, focused using a tube objective ($f_{TL} = 200mm$, MXA20696, Nikon) and captured using a sCMOS Camera (Zyla 4.2, Andor) after passing through the dichroic mirror.

In addition a separate 850 nm laser (0.9 mW, continuous wave circular beam, Thorlabs) was reflected from the coverslip by total internal refraction after passing through a clean-up filter (LL01-852, Semrock). The position of the reflected beam was determined by a linear light sensor (TSL1401CL, AMS-TAOS USA, Inc.) after passing through a 850nm band-pass filter (86-090, Edmund Optics). This was used as an input for the pgFocus open hardware focus module [15], which allowed to lock the position of the microscopes stage within a standard derivation of 10nm.

The images were acquired using the open source μ Manager [12] software. After selecting a field of view and locking the auto focus, the shutter was opened exposing the sample to laser illumination. When the emission density was low enough for commencing the STORM imaging, 3000 images with an exposure time of 10 ms were captured. To increase image quality for the second half of the image acquisition process the shutter of the 405 nm laser was opened after the fluorophore emission density became too low. This process was repeated until a sufficient number of field of views were captured.

The reconstruction of a single super-resolution image from this image stack is described in section 4.4.

4.3.3 SIM Sample preparation

The wild type and mutant stains were electroporated with mCherry-MTS2 plasmid to stain the inner cell membrane.

Cells were grown overnight and cell synchronization was performed as described in section 4.3.1.

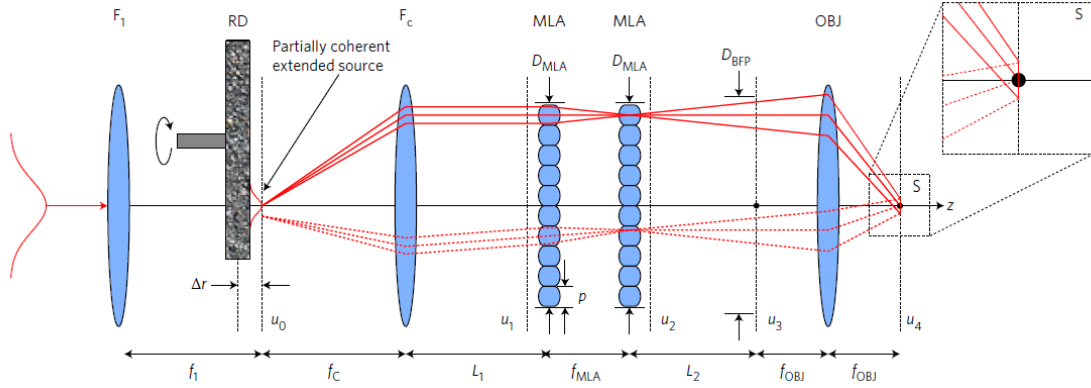


Figure 23: Schema of the flat-field illumination. The input Gaussian beam is extended by the rotating diffuser (RD), with the extension size depending on the offset δr . After passing F_c the light is splitted into independent ray bundles by the microlens arrays (MLA). These bundles are then redirected by the objective (OBJ) ensuring overlap in the sample plane. The illumination light comes from all points and a range of solid angles in the source plane, thus creating flat-field illumination (solid vs dashed lines). From [11]

For imaging, a silicon gasket filled with 1% M2G agrose was placed onto a cover slip, sealed using top cover slip and rested for 5 min. Afterwards, the top cover slip was removed and $1\mu\text{l}$ of bacteria suspension was pipetted onto the pad, keeping the pipette tip from touching the pad. To ensure aerobic growth conditions a small stretch of agarose was cut from opposing sides. After the drop was adsorbed to the pad, the gasket was sealed with a plasma cleaned round cover slip. During microscopy the temperature was kept at 28°C , allowing the cells to grow.

The full protocol including preparations for additional measurements is described in [2].

4.3.4 SIM Microscopy

Imaging was performed on a commercial SIM microscope (3D NSIM Nikon with 100x, NA 1.49 CFI Apochromat TIRF objective), equipped with 480 nm (480 mW) and 561 nm (400 mW) lasers and a back-illuminated EMCCD camera (512x512 CCD, iXon 3, Andor Technology).

Imaging of the mCherry-MTS2 fluorophore was performed on the 561 nm channel. The camera was operated at the maximum readout speed (1 Mhz) and dynamic range (16 bit). Preamplifier and electron multiplication gain were set to 1 and 200 respectively, maximizing the signal to noise ratio. All images were obtained with a acquisition rate of 200 ms and a laser power of $4\text{W}/\text{cm}^2$ balancing image quality and photo-bleaching. The microscope was set to 3D image mode for maximum signal to noise ratio and lateral resolution.

At total of 15 images were captured for each field of view ($30.7\mu\text{m}\times 30.7\mu\text{m}$) with 5 phase-shifted images at each of the 3 interference pattern angles. The acquisition of a full image stalk for reconstruction the super-resolution image took about 17s.

Time-lapse imaging was used to obtain cell dynamics. An image was captured ever 5 minutes, allowing to follow cell dynamics while minimizing the sample photo-bleaching due laser irradiation. Sequential imaging of multiple fields of view at each timepoint allowed to

follow up to 200 cells per experiment.

The reconstruction of the raw images was performed using the Nikon NIS-Elements software.

4.4 Image analysis

The first step for the analysis of STORM images was the conversion of the image stack recorded by the microscope to localization data. The conversion was performed by custom matlab code [11] producing a csv file containing precision and accuracy information for all localized molecules. The csv file was then rendered by the thunderstorm plugin of the ImageJ software [28], producing the bitmapped images required by further steps of the pipeline.

These images were subsequently segmented and analyzed for various shape parameters using a newly written matlab program. The results for each image were exported as csv files. An analysis of cell dynamics was then performed using a specialized ipython notebook. [30].

An overview of the complete image analysis pipeline is also shown in figure 24.

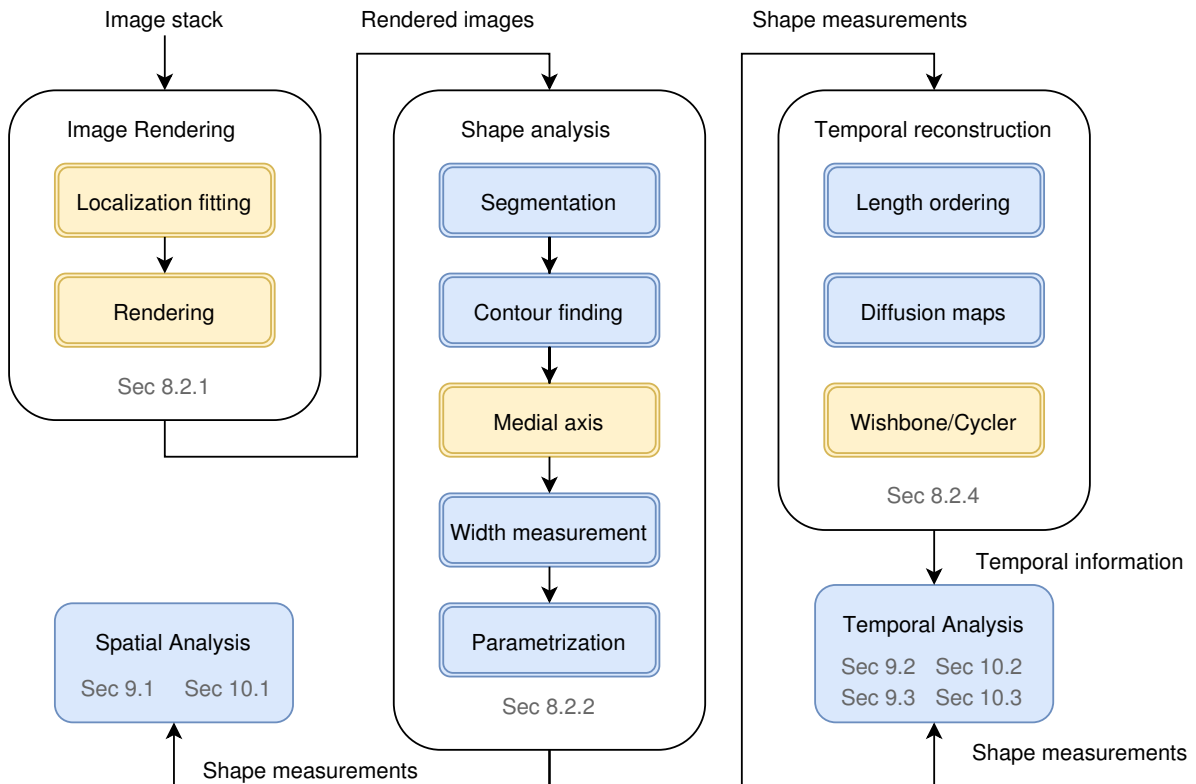


Figure 24: Graphical overview of the image analysis workflow. The arrows represent data dependencies and the colors indicate if the code is mainly based on external libraries (yellow) or has been developed specifically for this thesis (blue). Each process is described in more detail in the referenced sections (gray text).

4.4.1 Localisation rendering

Fitting

After the cells were prepared and imaged as described in section 4.3.1, the localizations for each light-emitting dye molecule needed to be computed for each microscopy image. This was performed by a customized matlab program, which performed a Gaussian fit on the emitters of each image. The localized emitter positions together with their uncertainties were then saved to a csv file.

Since the emitter density on the first 500 images was typically too high for reliable single emitter fitting, these images were discarded before the analysis. The calculations on a large number of images required substantial computational resources and were performed on a workstation's graphically processing unit (GPU).

Rendering

The resulting localization file was then imported into the ImageJ plugin ThunderSTORM [29]. As a first step the plugin was used to perform a drift correction, where the correlation between two sequential images was used to estimate the drift between two frames. Afterwards the image was rendered by assigning a Gaussian function to each localized point.

4.4.2 Shape

Overview

The analysis of the rendered images was performed in three steps. First, the field of views containing multiple bacteria were segmented into images of individual bacterias. Second, the length along the medial axis and width perpendicular to it were computed for each bacteria. And finally, the bacterias volume and surface area were obtained by locally rotating the bacteria around its medial axis. An overview of the measurement steps is shown in figure 25.

Segmentation

As a first step, the input image containing multiple cells was split into multiple single cell images. A connected components analysis was used to determine regions of interest. Only regions above a minimum pixel area threshold were selected to filter out background noise.

Afterwards, the regions of interest were cast into rectangles and enlarged by a few border pixels to account for possible outliers. These regions of interest were then used to carve out the individual cell images. The individual cells were reviewed manually in order to filter out damaged cells.

Contour finding

After the segmentation mask separating the bacteria from the image background was obtained, the next step was to convert mask to a contour. This was done by following the border of the masked region using the `bwtraceboundary` matlab function. The resulting list of boundary pixels represents the bacteria outline.

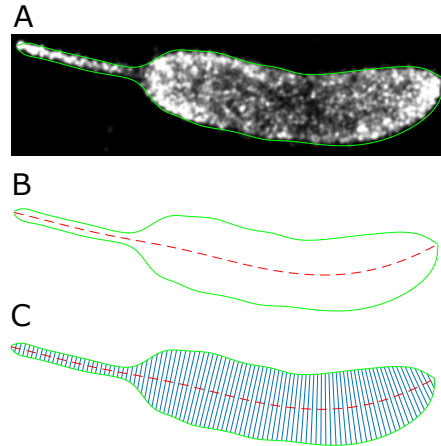


Figure 25: Overview of the measurement steps for a single bacteria. (A) The contour is extracted from the image (green line). (B) The medial axis (dashed red line) is computed based on the contour. (C) The width is measured at images slices locally perpendicular to the media axis (blue lines).

Medial axis

After the bacteria outlines were obtained, the medial axis was computed in order to straighten the bacteria for further measurement. The medial axis was approximated by computing a Voronoi diagram of the bacteria outline and removing all except the main branch.

For this computation the MicrobeTracker library [37] was used. The obtained medial axis was then smoothed using a spline to provide a good basis for further measurements. Using the same library the cell contour was also split into two parts, one above and one below the medial axis called mesh.

Width measurement

Cell width was measured relative to the medial axis. The measurement was performed by analyzing slices of the cell image locally perpendicular to the medial axis. The image obtained by appending these slices is shown in figure 26. Two methods were used to determine cell width from these slices. The widths could either be computed by measuring the FWHM of the slice or by taking the width of the cell mesh at each medial axis point. The following section describes both approaches in more detail.

FWHM

By computing the FWHM along each slice the width could be measured well for most of the cell. However, the measurements start to significantly deviate from the cell shape for regions of small width. This can be caused by bias within the FWHM measurement due to non-uniform fluorescent labeling of the cell or by the effects of the spline based smoothing. Both effects are shown in figure 27a.

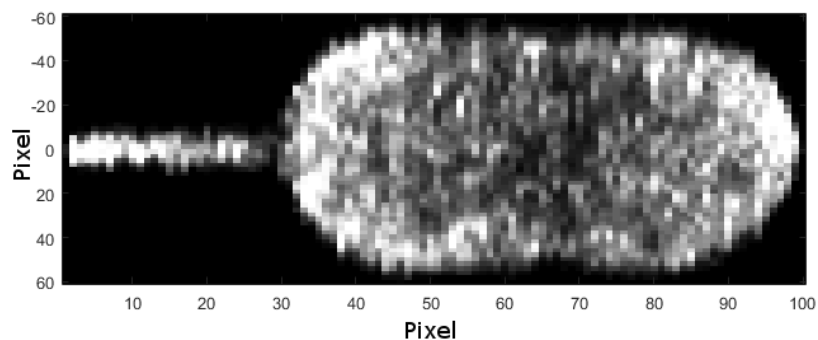


Figure 26: Image of bacteria straightened along the medial axis. The X axis corresponds to length steps of equal distance along the medial axis while the Y axis represents the image slices locally perpendicular at each measurement point. The Y axis magnitude represents the distance of the image slice points from the medial axis.

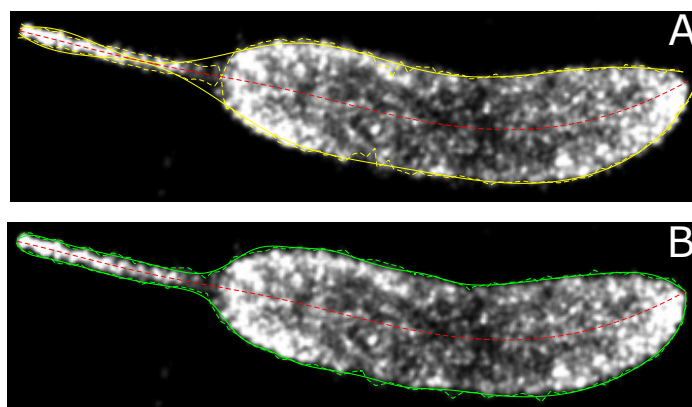


Figure 27: Comparison of contour measurement methods. Dotted lines represent the measured, solid lines the smoothed contour. The dotted red line corresponds to the cell medial axis in both images. (A) Contour measurement using the FWHM. (B) Contour measurement using the cell outline (mesh).

Mesh

The width measurement based on the mesh was computed by calculating the distance between corresponding contour points above and below the medial axis. Since it is solely based on the cell outline it tends to overestimate the true cell width. For the same reason it is also more robust with respect to non-uniform fluorescent labeling of the cell. This is also shown in figure 27b.

Stalked and swarmer cells

The *C. crescentus* cells used in this project featured two compartments with different properties. One compartment builds a stalk and is therefore called the "stalked" cell while the other builds a flagella and is called the "swarmer" cell.

These structures are much thinner than most of the cell and differ significantly in their properties. It is therefore desirable to separate the measurements done on the stalk from the ones performed on the rest of the cell. In this project the separation was performed by searching for the end of the flat region corresponding to the stalk. To separate stalked from swarmer cells it was assumed that stalk should have a length at least 15% of the total cell length.

DNA

For the cells with DNA staining as described in section 4.3.1, the DNA content needed to be quantified. Since for the spatial resolution gains were not significant for DNA images, diffraction limited-whitefield images were used for the quantification. This simplified the analysis pipeline, as the images could be captured in dual color using a different wavelength to excite the DNA dye.

The amount of DNA content was estimated by assuming a linear proportionality with the integrated signal intensity of the DNA dye within the cell. This assumption stems from the fact that the used DNA dye is activated when intercalated within the DNA strand.

All images were taken with the same laser power and exposure settings, taking care to avoid signal saturation, to allow for a unbiased comparison between different field of view. In addition, a camera dark-field was captured and subtracted from all images prior to analysis to avoid a location dependent intensity bias.

To estimate the integrated intensity within the cells, cell outlines were required. These outlines were obtained by segmenting whitefield images of the membrane dye and transferring the regions to the DNA image with equal image dimensions. The signal intensity per pixel was then summed up and reported as DNA signal.

Shape parametrization

To parametrize the measured shape, the model shown in figure in figure 28 was used. In order to fit the model efficiently, the fit was performed stepwise. First, the previously calculated stalk length was used to separate the cell into stalked (l_s) and cell body part (l_c). Then the maximum diameters within the stalk and cell body were used as r_s and r_c respectively. The minimum width within the cell body excluding the poles was used as r_{min} and its position determined l_{div} .

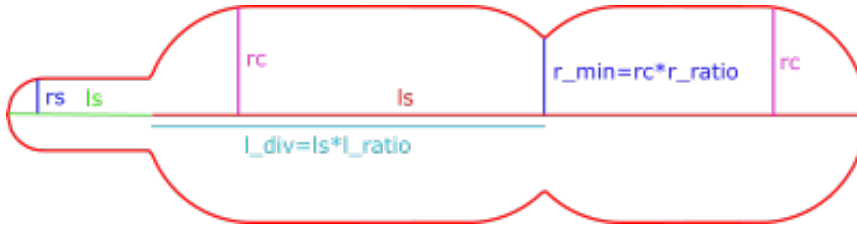


Figure 28: Shape parametrization used for *C. crescentus*. The method of parameter extraction from the measured shapes is described in the text.

4.4.3 Feature importance

Using the previously described steps, a number of features describing the bacteria were obtained. From the raw data it is not immediately clear which set of features would be best suited for the temporal reconstruction of cell cycle time. A method to find an optimal set of features would therefore be desirable. Such a method could use the time measurement SIM to assess the importance of the individual features.

A method fulfilling these criteria is the feature importance from random forest regression. Each decision tree in the forest can compute the importance of individual features by looking at the variance induced by the feature. The average over all trees in the forest then represents a measure for feature importance. In addition, the standard deviation allows to estimate the variability of feature. This method was found to perform well on biological data [3].

For implementation, the `RandomForestRegressor` of the scikit-learn [34] library was used in this thesis. The number of trees was set to 30 with 30% of the SIM data used for training. In addition to computation on the full dataset, the feature importance was also estimated for bins containing 20% of the cell cycle to find a possible change of feature importance over time.

4.4.4 Temporal reconstruction

The temporal reconstruction was implemented using three different methods. Since these methods should be able to reconstruct time without previous calibration, only the unsupervised methods from section 3 were considered.

Length ordering

The implementation of length ordering was rather straight forward. All cells were ranked based on their length excluding the cell stalk from smallest to largest. Each cell was assigned a time linearly interpolated between 0 and 1 according to its rank. The ERA correction described in section 2.3 was applied and saved separately from length ordering time.

Diffusion maps

Following [18] a python implementation of diffusion maps was developed for this thesis. From the analysis of feature importance cell length and waist ratio emerged as the most important features and were therefore used as input. Both features were normalized separately between

0 and 1 before passing them to the diffusion maps algorithm. This prevents differences of the measurement scale from influencing the diffusion map performance.

Apart from the data, the diffusion maps algorithm requires a measure for the Gaussian kernel width (sigma). The Gaussian kernel width was determined using the method of average dimensionality peak described in [18].

The application of this method resulted in a diffusion vector for each measured bacteria. The entries of the first component of the diffusion vector represent a measure of similarity between cells. The extracted components were then normalized between 0 and 1 to yield the diffusion time. Diffusion maps has no information on the direction of bacteria dynamics. Therefore, the assigned diffusion time was reversed if the last cell was found to be smaller than the first cell.

In addition, a ranked diffusion time as calculated which assigned assigned a time between 0 and 1 base on the diffusion time rank.

Wishbone

The usage of wishbone [35] required computing the diffusion maps as described in the previous section beforehand. In addition wishbone takes to additional parameters: a cell used as starting cell and the number of diffusion map components to consider. The number of components was set to 2, since only measurements per cell were provided. As a starting cell, the smallest cell was chosen. In addition wishbone allows to restrict the calculation to the nearest neighbours. A value of 30 was used as a compromise between calculation speed and accuracy. As described in section 5.3 the results of wishbone on the datasets used in this thesis were rather erratic and therefore not investigated further.

5 Results

To address the questions raised in the previous sections, the analysis of the experimental results is divided into 3 sections. The first section assesses the influence of spatial resolution differences between the microscopy methods and the trade-offs of shape parametrization. In the next section, the impact of the ability to follow single cell trajectories is discussed. And lastly, the final section compares the developmental trajectories obtained using pseudo-temporal classification to the measured trajectories.

For easier reference, this partitions is also maintained in the discussion section.

5.1 Spatial resolution

This section compares the measurement of the cell shape obtained using various microscopy techniques as well as different representations of shape.

High spatial resolution is essential for accurate tracking of cell division

When comparing the cell profiles obtained with diffraction limited to the ones obtained with super-resolution techniques it can be seen that lower spatial resolution results in an over-estimation of cell width and length (Figure 29). For PALM and phase-contrast images of

the same bacteria were available while for SIM a cell with similar length and developmental stadium was chosen.

The PALM and SIM images imaged the fluorophores located at the inner membrane while the phase contrast images observed the outer shape. Part of the difference may therefore be due the membrane thickness of the cell. This difference is further discussed in section 6.1.

Since the absolute values of cell width are smaller than cell length, the percentage uncertainty due to spatial resolution is higher for width. This limits the ability to accurately track cell division towards the end of the cell cycle since percentage uncertainty increases for small widths observed at the division site.

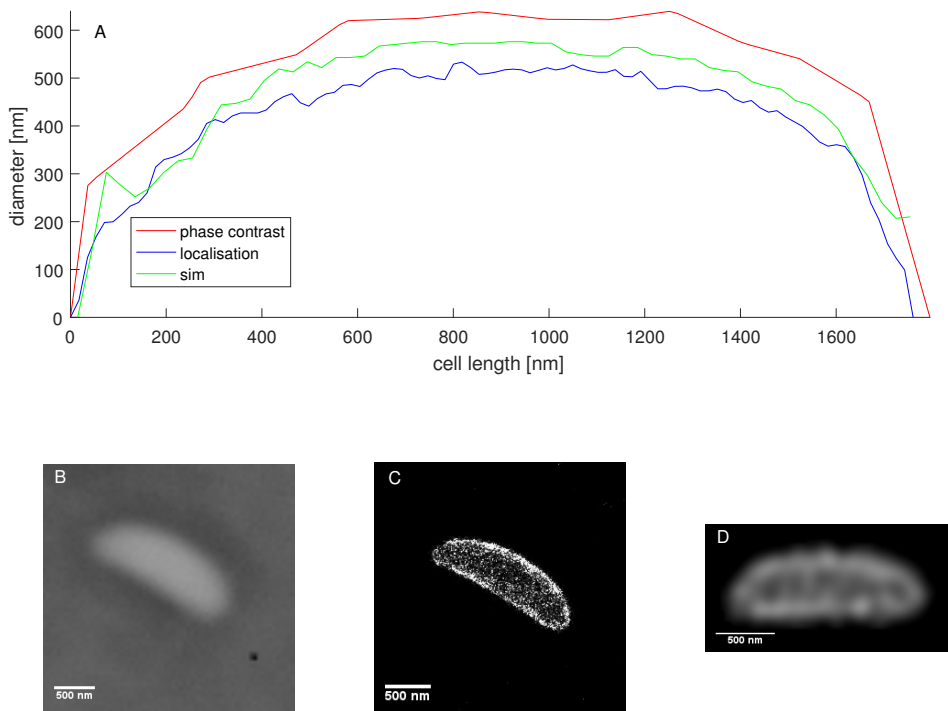


Figure 29: Comparison of cell width measurements using phase-contrast and localization microscopy on similar cells. (A) Cell width measurement as a function of length along the medial axis. A strong discrepancy in width measurement and a small discrepancy in length measurement can be observed. (B) Phase contrast image used for the measurements (C) Localization image of the same cell. (D) SIM image of a different cell with similar length and developmental stadium.

Parametrization can reduce noise at the expense of cell variability

Cell parametrization can reduce the measurement noise at the expense of a simpler overall shape. The classical model for *C. crescentus* assumes a cylindrical body with hemispherical caps [19]. Fitting this model to actual measured shapes shows that it provides a good first approximation for the measured cell shape.

However, the exact functional form differs from cell to cell. While some invagination sites closely reassemble capped hemispheres, others show more a complex functional form. This is exemplified for two cells measured using STORM in figure 30.

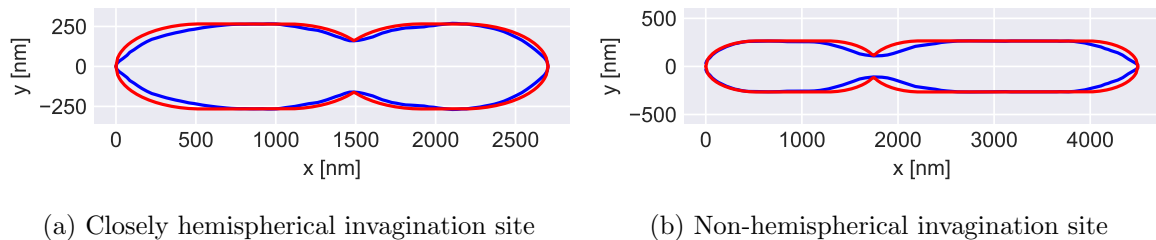


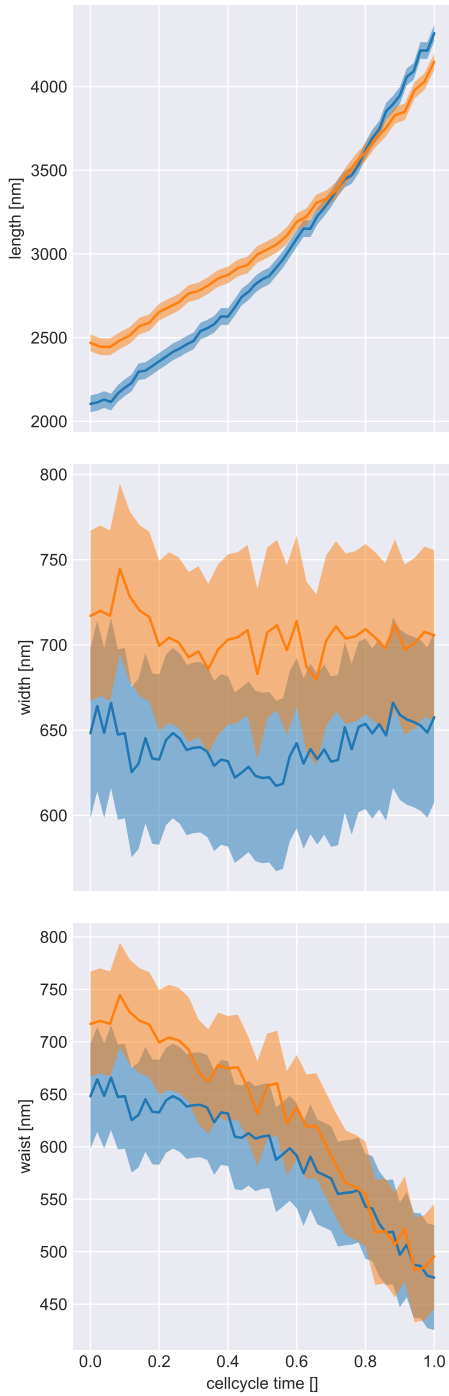
Figure 30: Comparison of shape parametrization for two invaginating cells measured using STORM. The measured contour (blue) is shown together with the fit based on the hemispherical model (red). While for the cell in figure 30a the capped hemisphere model closely reassembles the measured shape at the invagination site the invagination site in figure 30b is elongated in comparison with the hemispherical model

5.2 Single cell identification

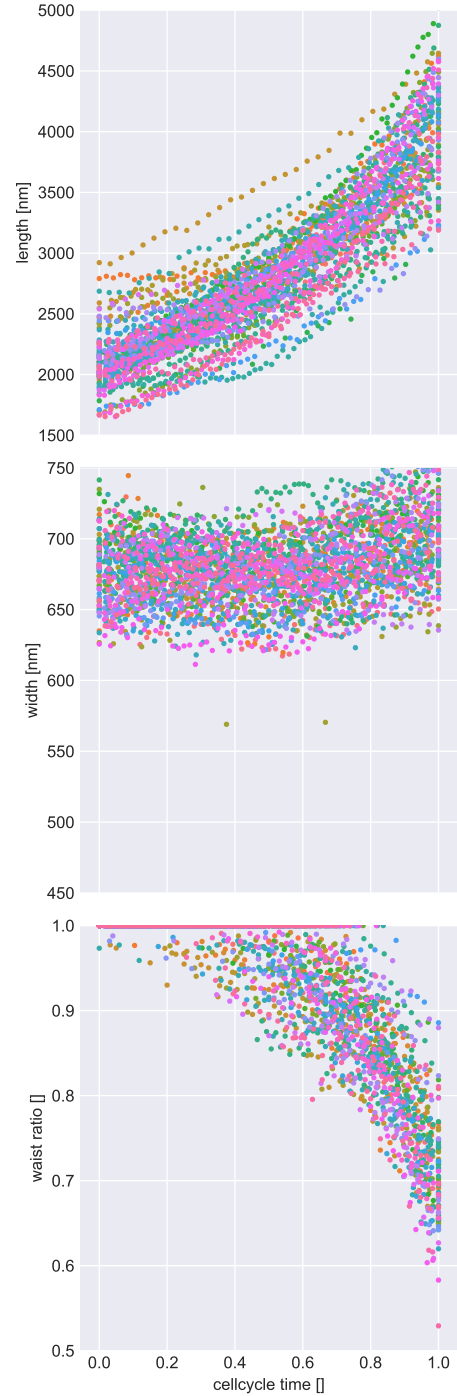
Time-lapse SIM microscopy allowed to follow the development of single cells. This section assesses the differences between single cell and population trajectories. In addition the results of methods to compute an average contour from single cell shape measurements are discussed.

Population noise largely obscures single cell trajectories

Populations exhibit a wide variability of shapes. While single cell observation show a clear developmental trajectory over the course of a cell cycle, this development is largely obscured in population measurements. For example the two single cell trajectories shown in Figure 31a follow different rates for length and waist development, something which can not easily be inferred from the population measurement of figure 31b. Depending on the required accuracy of the study, single cell identification may therefore be crucial to obtain precise developmental trajectories.



(a) Single cell trajectories



(b) Population trajectories

Figure 31: Comparison of 2 selected single cell trajectories to the population measurements obtained using SIM. (a) The measured trajectory of two bacteria is represented by a thick line while the bands correspond to the uncertainty induced by the measurement resolution. (b) Scatter of population measurements where each single cell trajectory can be identified by the assigned colors ($N = 3238$)

Combining bacteria trajectories requires sound normalization

Within a population there is a significant variation in cell contour measurements. To compute an representative contour for the overall population it is therefore essential to normalize the cells first, allowing a combination of cells irrespective of differences in absolute measured values.

For the *C. crescentus* populations analyzed in this thesis the normalization has to allow comparison between the shapes of the stalked and swarmer cells during cell division. Since the stalk has a growth behaviour different from the rest of the cell it is treated separately from the cell body.

In addition, cell invagination may occur at different positions along the cell body. To ensure that the invagination is kept when averaging over multiple bacteria, the cell body is normalized separately for the parts left and right of the invagination site. For cells without measured invagination the invagination site was assumed to be in the middle of the cell body.

After this normalization has been performed a method for combining the resulting contours needs to be chosen. From the methods described in section 4.2, a contour based on lowess smoothing showed the best results.

Alternatively, it is also possible to average over the parameters of the parametrized shape instead of normalizing the contour measurements. This generally leads to a smoother, less detailed shape (figure 21).

5.3 Temporal reconstruction

Pseudo-temporal methods allowed to reconstruct developmental trajectories from images of fixed cells. This section describes conditions required to apply pseudo-temporal methods successfully and compares the results of the methods to measured developmental trajectories.

Length and waist ratio are the most influential parameters for pseudo-temporal reconstruction

Assessing how well cell cycle time can be estimated from cell shape by using machine learning (random forest) leads to the following conclusion: Good estimates are possible for the end of the cell cycle while classification performance decreases significantly for cells in the beginning of the cell cycle (figure 33). As can be seen from figure 32 the most significant shape parameters for random forest are cell length and to a lesser extend waist.

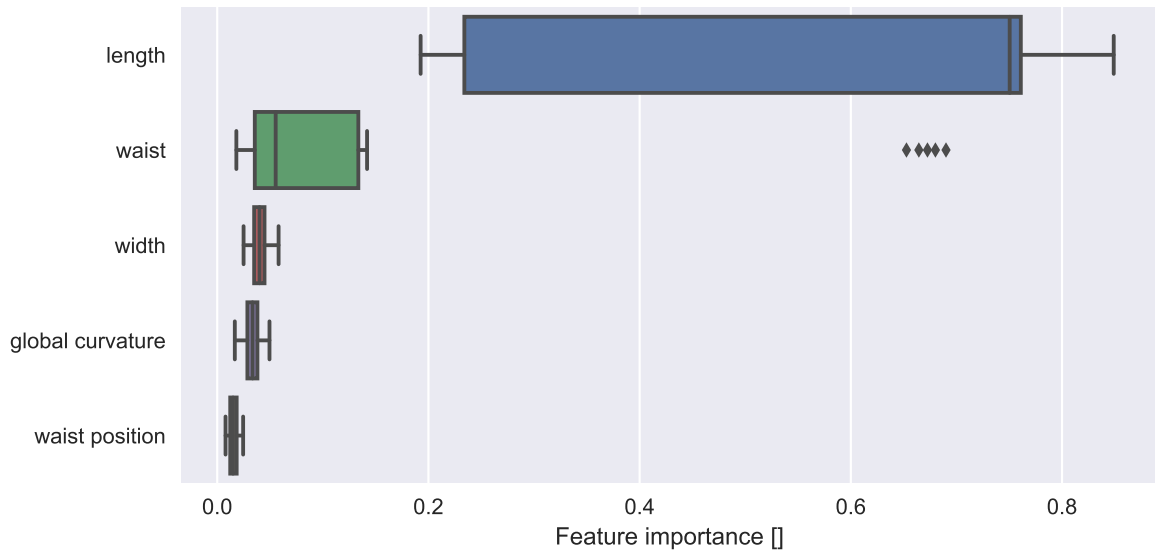


Figure 32: Comparison of shape parameter importance for the random forest regression. Length and waist stand out as the most influential features. The box represents the quartiles of each feature with the whiskers extending to the rest of the distribution cleaned of outliers. The outliers are represented as small diamonds, the medians by vertical lines within the box.

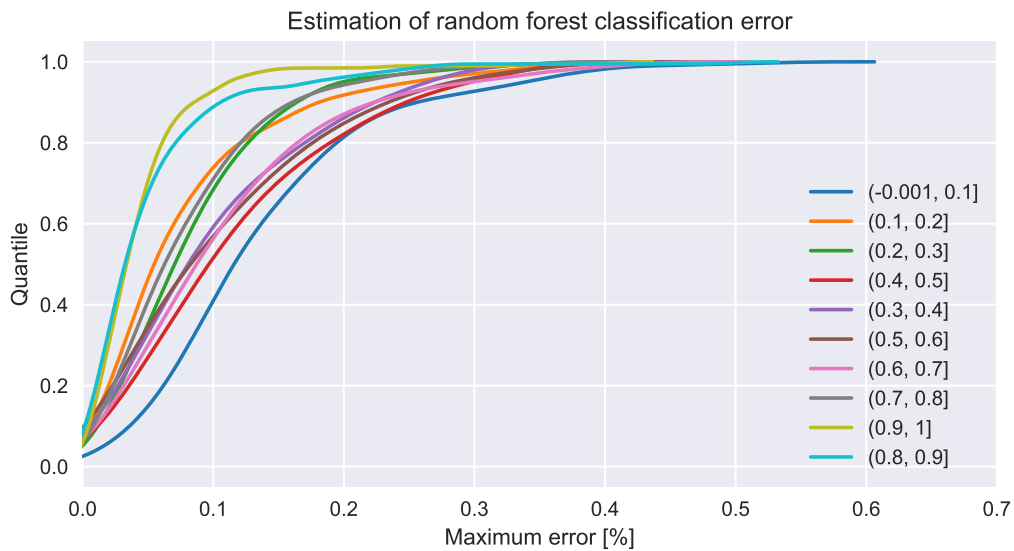


Figure 33: Estimation of classification error with random forest for 10 time bins. Each bin is represented by a different color. It shows the cdf of the classification error in each time bin. For example in the 0.8 to 0.9 time bin, more than 90% of the measurements were classified with less than 10% error.

The importance of waist surpasses length for the end of the cell cycle

Figure 32 suggests that length and waist are the most important features for temporal reconstruction. However, this analysis did not distinguish between phases of the cell cycle, where the feature importance may be different from the overall average. Repeating the analysis of figure 32 for length and waist for different fractions of the cell cycle shows that the importance of waist surpasses length towards the end of the cell cycle (figure 34). An increase in waist importance towards the end of the cell cycle is expected since cells start dividing about half-way through the cell cycle.

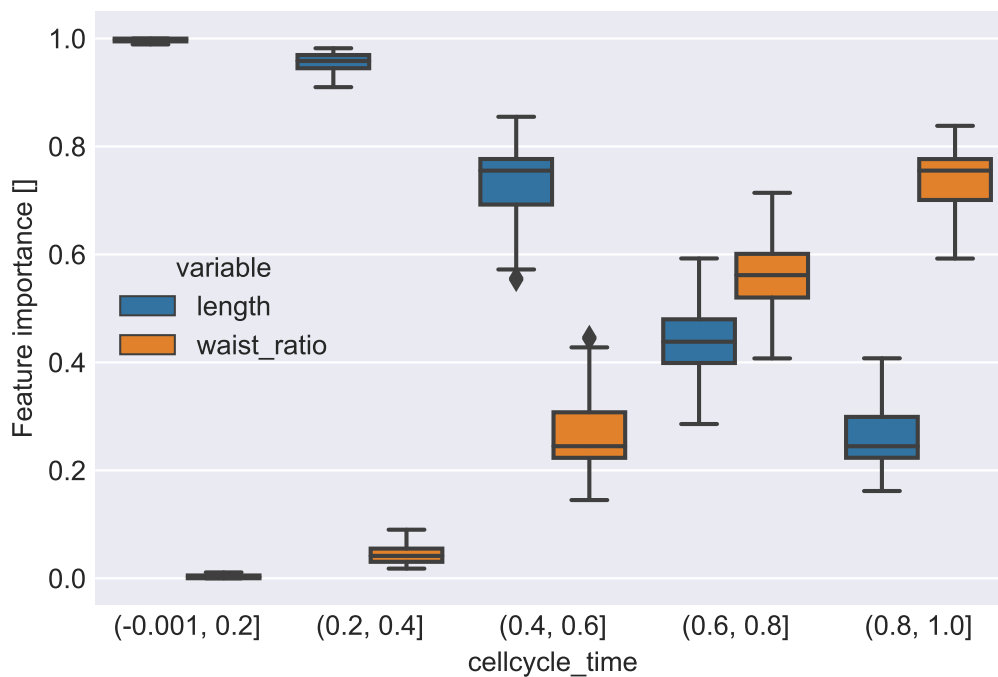


Figure 34: Comparison of feature importance estimated using random forest regression for different fractions of the cell cycle. The fractions of cell cycle time between 0 and 1 are given in brackets. The importance of waist increases during the cell cycle, surpassing length in the 60 to 80% time bin. The box represents the quartiles of each feature with the whiskers extending to the rest of the distribution cleaned of outliers. The outliers are represented as small diamonds, the medians by vertical lines within the box.

The calibration of supervised method is generally only valid within a single strain

Comparing the average obtained using the supervised methods described in section 3.1 to the averages obtained using measured time showed a clear limitation. While the average of figure 35 (calibrated on wild type data) was following the general trend of the measured averages for the wild type strain, this wasn't the case for the mutant strain. The mutants

length is shorter on average, therefore invalidating the calibration performed on the wild type. For nearest neighbour calibration the comparison between the averages yields similar results (figure 36). For new strains with unknown growth constants supervised methods may therefore yield unreliable results.

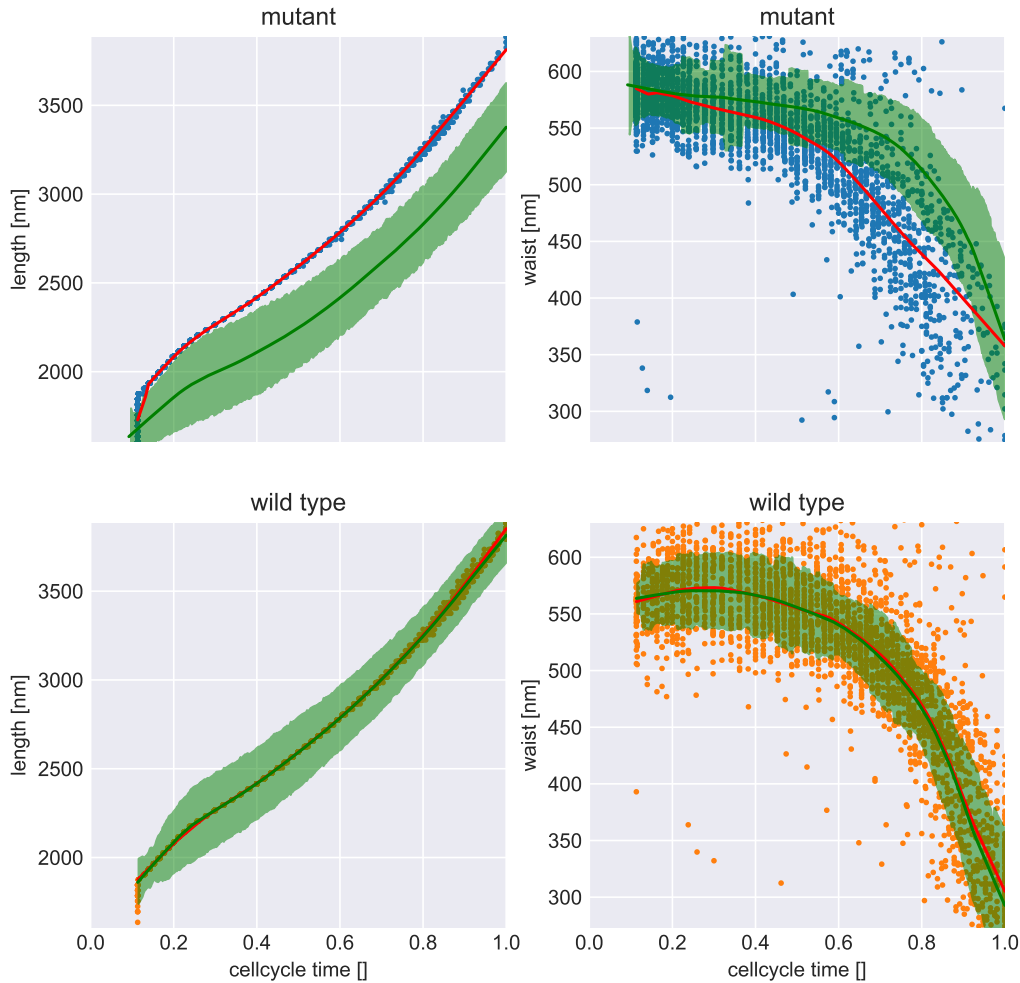


Figure 35: Comparison between the measured average (green line) and calibration curve (red line) population average (lowess, $\alpha = 0.3$) for wild type ($N = 2918$) and mutant ($N = 2927$) data. The calibration curve average was calibrated by the wild type measurements on all plots. Individual measurements (scatter points) were positioned using the time predicated by calibration. First measurements were taken 30 min after synchrony for the wild type and 35min for the mutant data. The shading represents the standard deviation of a sliding window containing 5% of the measured cells.

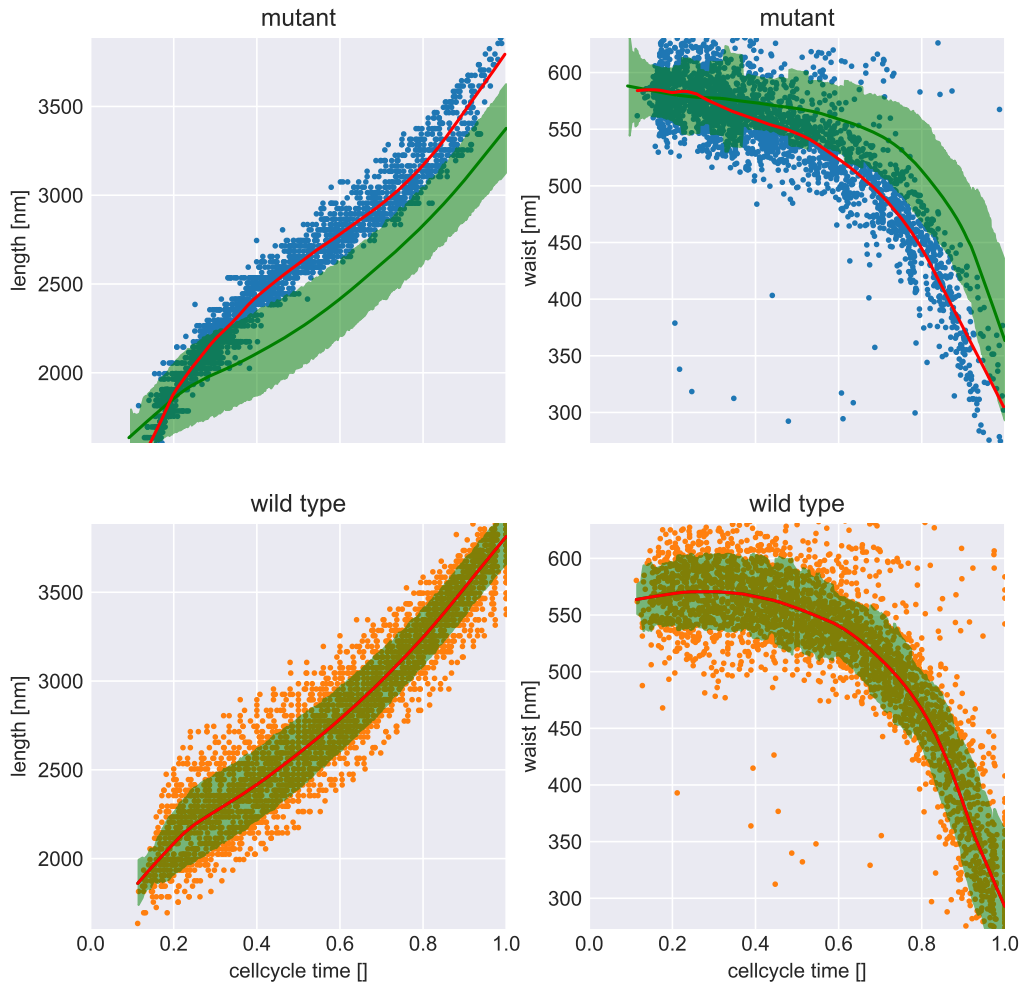
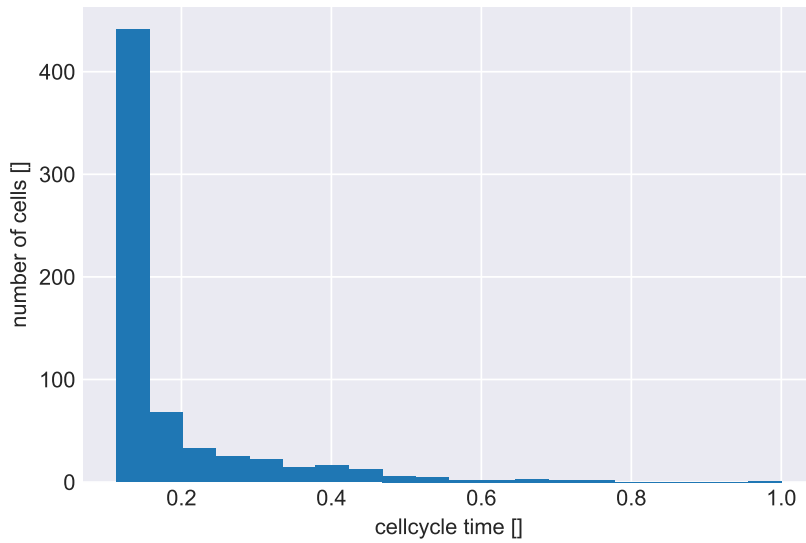


Figure 36: Comparison between the measured (green line) and nearest neighbour (red line) population average (lowess, $\alpha = 0.3$) for wild type ($N = 2918$) and mutant ($N = 2927$) data. The nearest neighbours average was calibrated by the wild type measurements on all plots. Individual measurements (scatter points) were positioned using the time predicated by calibration. First measurements were taken 30 min after synchrony for the wild type and 35min for the mutant data. The shading represents the standard deviation of a sliding window containing 5% of the measured cells.

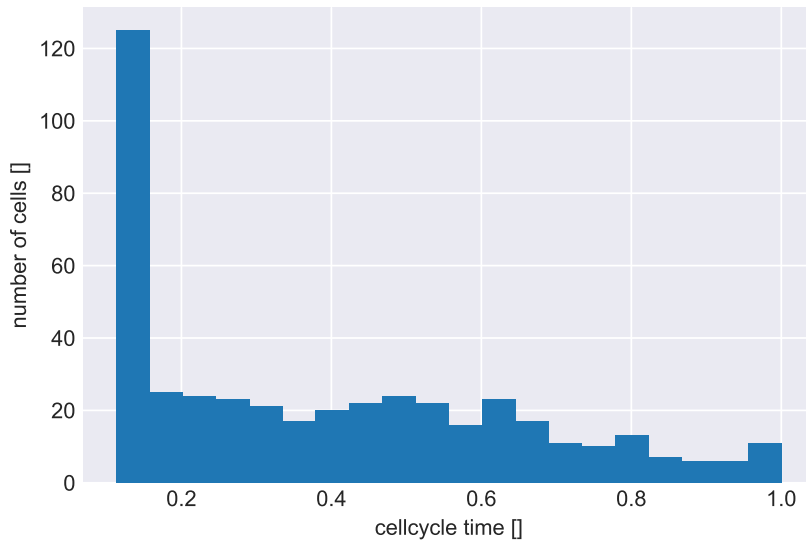
Calibration can be used to specify the observed portion of the cell cycle for the calibrated strain

After correcting the wild type STORM measurements for the resolution bias observed in the SIM measurements (section 4.1.1) their cell cycle times can be estimated by using a calibration curve for the wild type (section 3.1). Results for the synchronized STORM data indicate that primarily cells early in their cell cycle were observed (figure 37a), therefore invalidating the fair-sampling assumption required for length ordering classification. Length ordering of the unmodified synchronized STORM measurements may therefore yield misleading results. In contrast, the unsynchronized STORM measurements also contain a non-negligible number of observations in the second half of the cell cycle (fig 37b), indicating better sampling.

Both figures show a large number of cells in the first time bin. This behaviour is an artifact stemming from the missing calibration curve information for the beginning of the cell cycle (the measurement was started 30 minutes after synchrony). All earlier cells are grouped within the first bin, producing making it an outlier.



(a) Synchronized STORM measurements



(b) Unsynchronized STORM measurements

Figure 37: Time distribution of STORM cells as predicted by the SIM calibration curve for the same strain.(a) The majority of the observed cells come from the beginning of the cell cycle. Only very few cells are classified to be in second half of the cell cycle ($N = 654$). (b) For synchronized STORM measurements cells were also observed in the second half of the cell cycle ($N = 442$)

Ordering can provide good pseudo-temporal classification for most of the cell cycle

Comparing the different pseudo-temporal reconstruction techniques, each had their specific advantages and disadvantages. Length ordering performed best on the majority of the cell cycle (figure 38), only surpassed by diffusion maps for the last 15% of the cell cycle (figure 39). Wishbone struggled with the small number of parameters and produced erratic results. From a comparison of measurement trajectories using all 3 techniques (figure 42) it can be seen that the wishbone trajectories had little in common with the observed behaviour. Overall, the general population development can be represented surprisingly well using simple length ordering (figure 41).

Comparison between of the SIM and STORM averages for different populations of the same strain show that averages are similar even for different measurements (figure 40). In addition, visual comparison between the SIM and STORM images for the wild type measurements shows that the ERA ordered STORM images show a similar time dependant behaviour (fig 43).

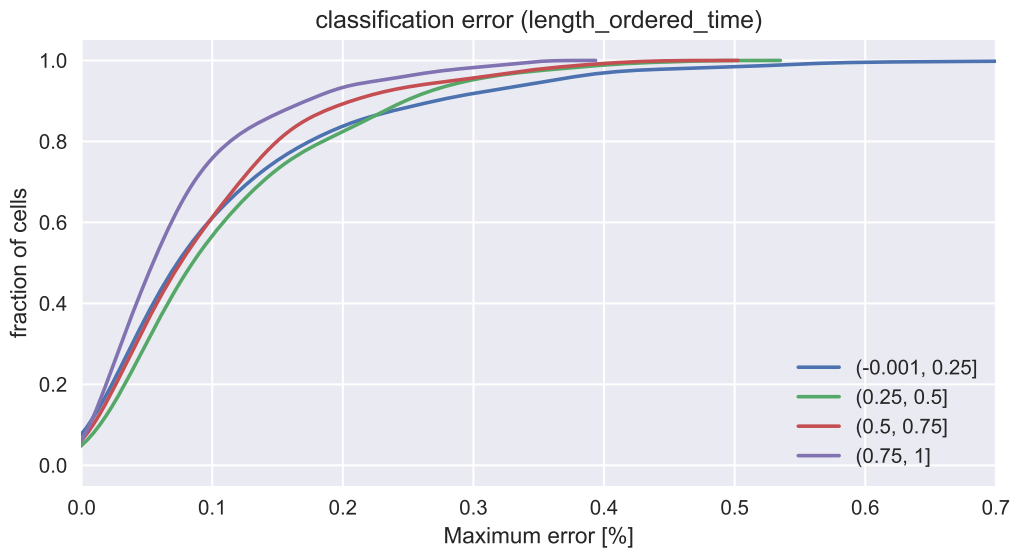


Figure 38: Pseudo-temporal classification error using length ordering for wild type data.

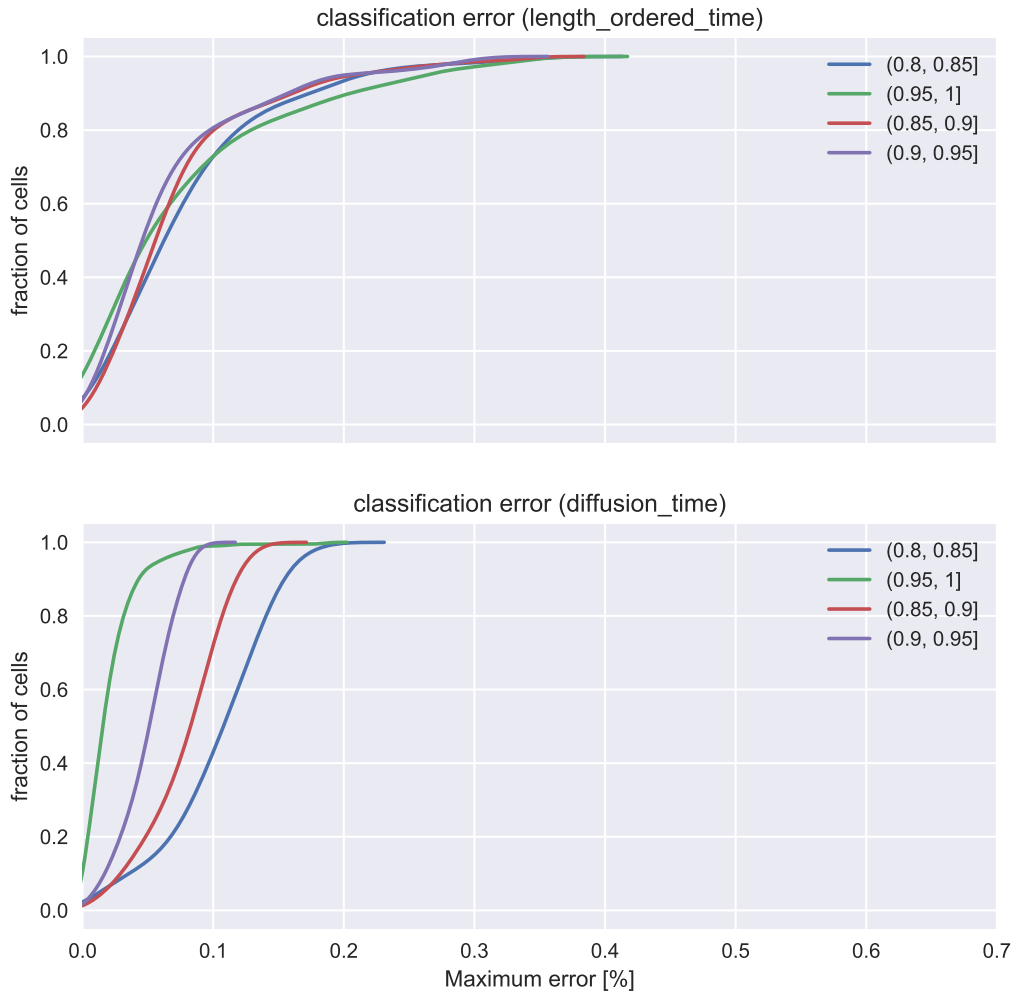


Figure 39: Pseudo-temporal classification error for the last 20% of the cell cycle using using length ordering and diffusion maps on wild type data. For the last 15%, diffusion maps performs better than simple length ordering (lower maximum error for a given fraction of all cells). Otherwise simple length ordering performs better (see figure 38).

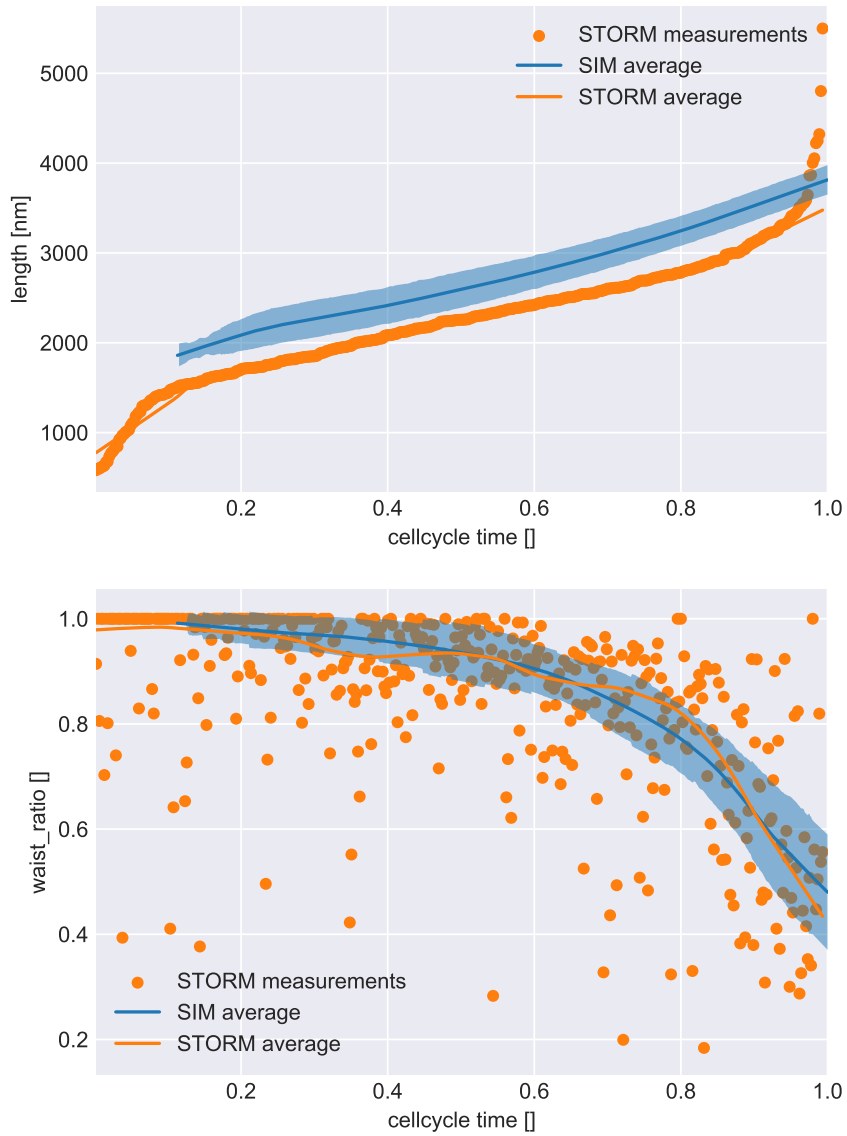


Figure 40: Comparison of population averages for different populations of the same stain using SIM ($N = 2918$) and STORM ($N = 442$) measurements. The shading represents the standard deviation of a sliding window containing 10% of the measured cells.

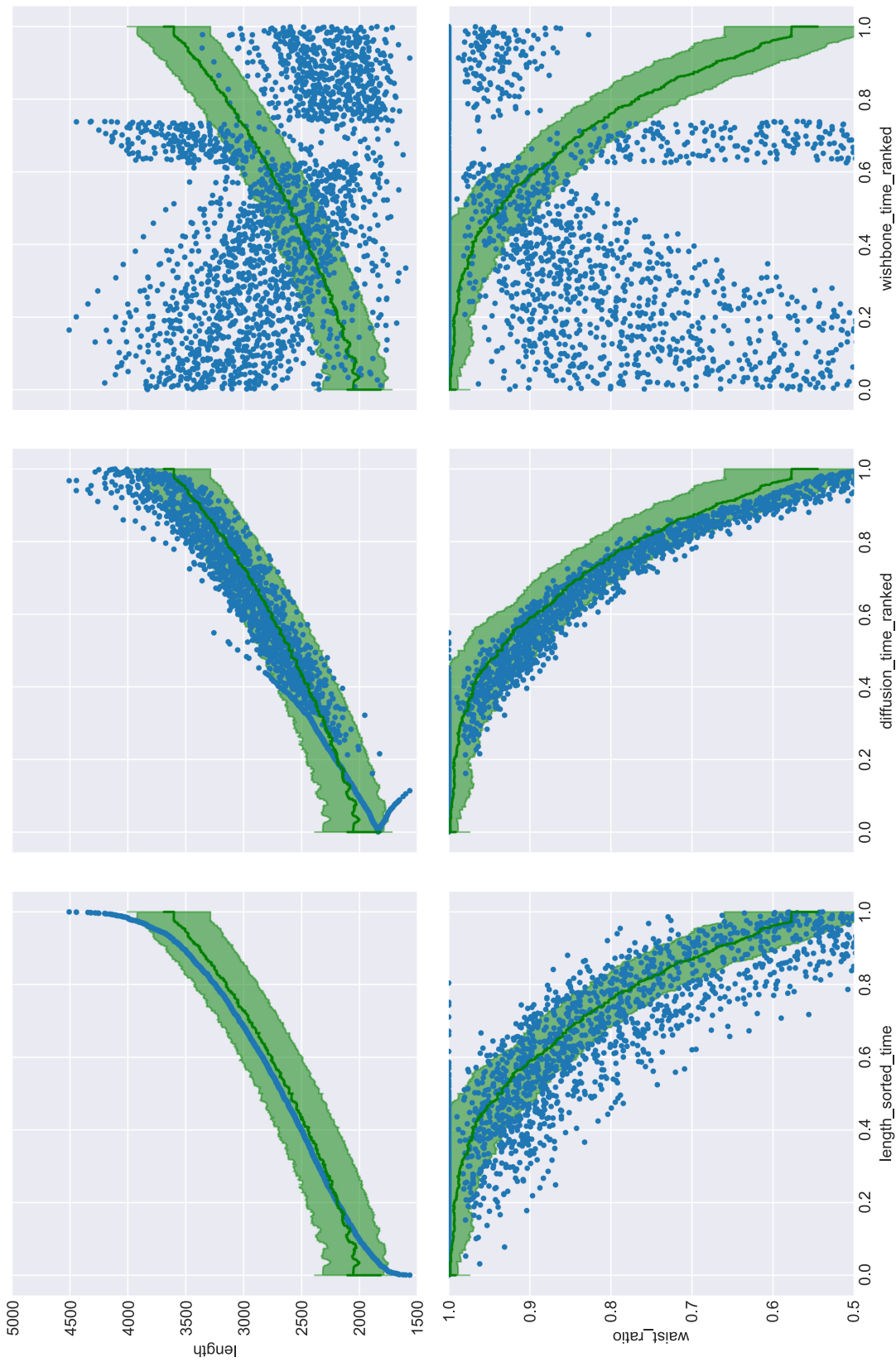


Figure 42: Comparison of measurement trajectories using different pseudo-temporal ordering techniques. The data is positioned on the time axis using pseudo-temporal ordering (blue scatter) and a comparison to the sliding average of measured time (green line) and standard deviation (green shading) is shown ($N = 2203, frac = 0.05$). It can be seen that diffusion time leads to a less noisy waist_ratio trajectory and contains spurious branches in the first half of the length trajectory. The trajectories based on wishbone time are erratic and just shown here for a full comparison.

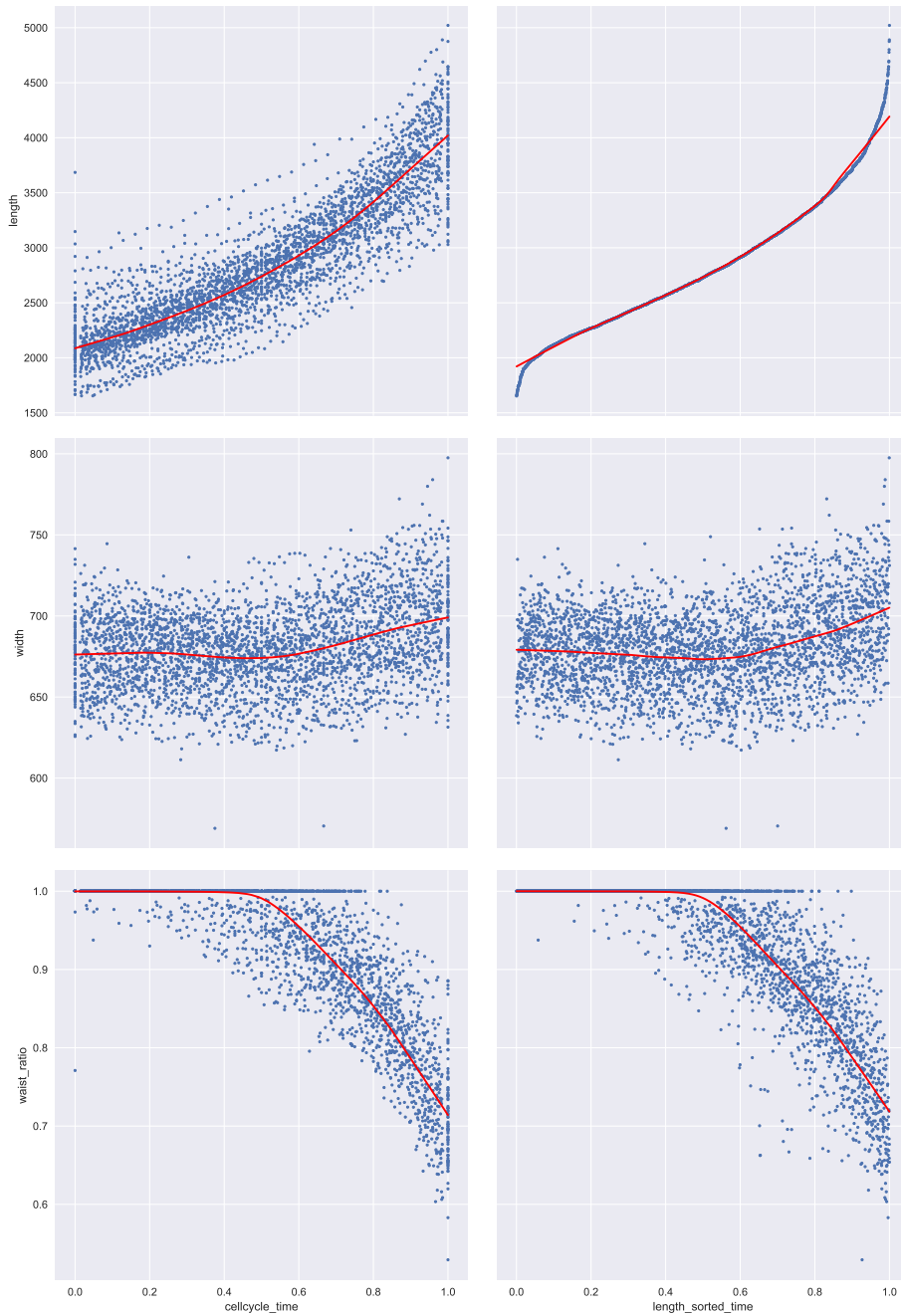


Figure 41: Comparison of population trajectories resulting from measured SIM time and pseudo-temporal length ordering ($N = 2927$). The blue scatter correspond to the individual measurements while the red line is a population average trajectory computed by lowess smoothing ($\alpha = 0.3$). The population trajectories are similar for width and waist ratio while for length outliers lead to the effects described in section 3.2.1.

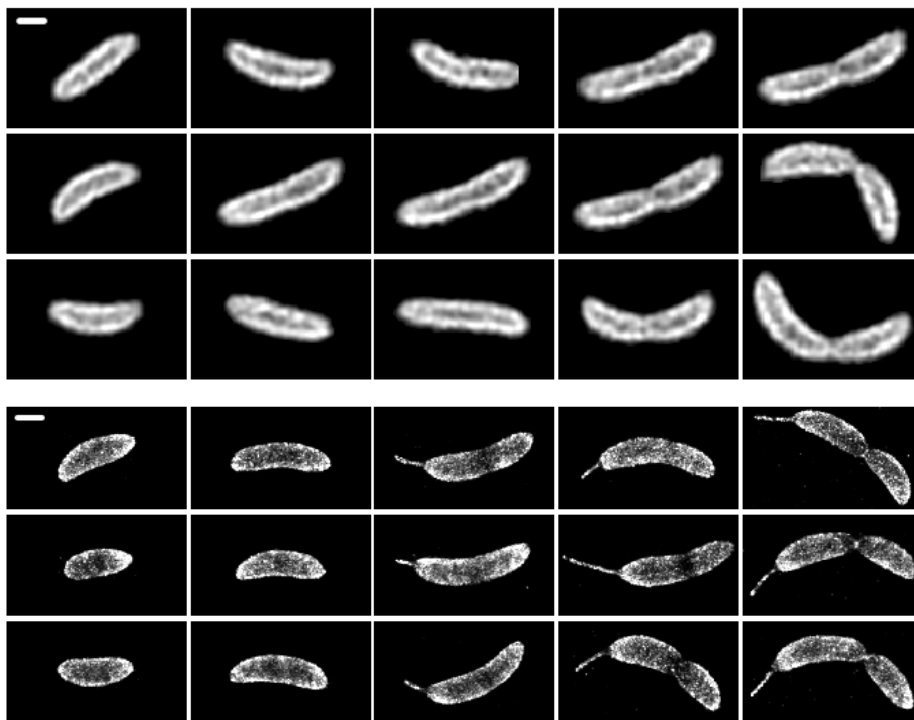
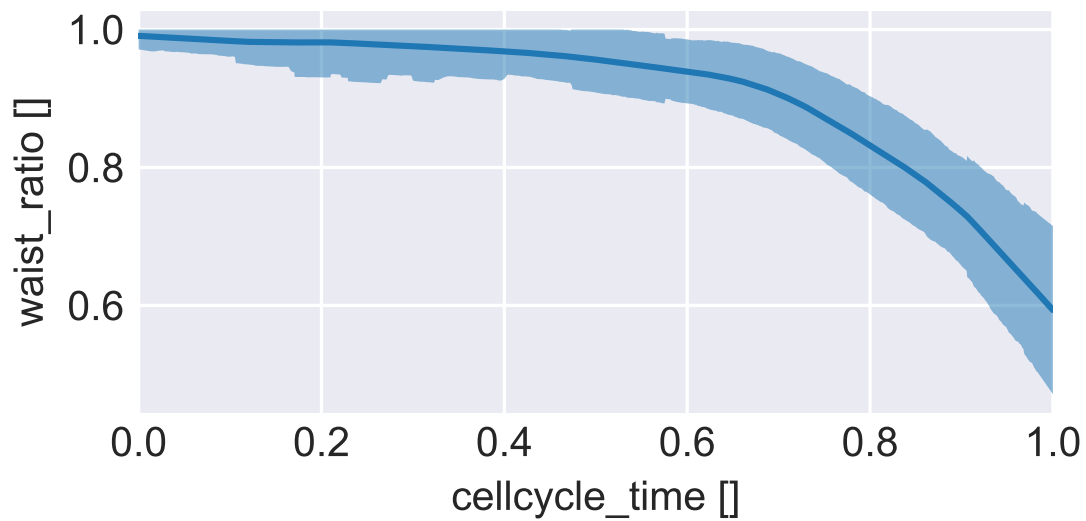


Figure 43: Comparison of selected SIM (top block) and STORM (bottom block) images for 5 time bins. The top graph represents the waist ratio average (lowess $\alpha = 0.3$) over the SIM cell cycle time measurements ($N = 2977$). The vertical dividers in the graph mark the time bin borders. Blue shading is the standard deviation of a sliding window over 10% of all cells. Scale bars 500 nm.

DNA measurements didn't improve pseudo-temporal classification for the available measurements

To check the benefits additional measurements for pseudo temporal reconstruction, DNA content measurements for each cell were added to the dataset. However, the temporal classification performance did not improve as theoretically expected after adding the DNA signal.

Possible reasons include differences in DNA development between stalked and swarmer cells, segmentation noise due to alignment differences between shape and DNA signal and intensity differences between the frames.

6 Discussion

6.1 Spatial resolution

Comparison between the diffraction limited phase-contrast and super-resolution images of *C. crescentus* has shown that phase-contrast images tend to overestimate the cell size (figure 29). This systematic bias may stem from a number of reasons. First, the phase-contrast images measure the outer membrane while the dye used for STORM and SIM images targets the inner membrane. Since the inner and outer membrane are typically about 50 nm apart a difference between the observed cell sizes is expected (figure 44). Second, the threshold based image segmentation will lead to a larger apparent cell size for lower resolution because the point-spread function fades slower, yielding a larger diameter for a fixed intensity threshold. While the systematic bias can be partially corrected, important features of *C. crescentus* remain hard to quantify using diffraction-limited microscopy.

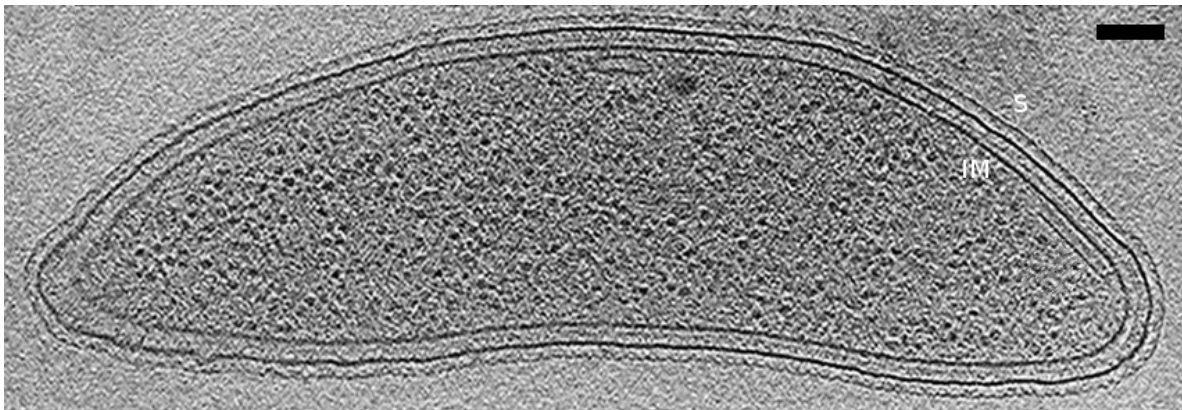


Figure 44: Cryo EM image of a *C. crescentus* cell without stalk. The inner membrane (IM) and surface (S) is separated by a distance of about 50 nm. The scale bar is 100 nm. Adapted from [6].

Such a feature is the invagination depth (waist) and position along the cell. Even when using super-resolution microscopy, a waist below 300 nm is hard to observe (figure 36). However, this information is crucial for the cell dynamics as constriction rate and duration may play a critical role in cell size control and homeostasis [2]. In practise the constriction rate can be extrapolated to obtain a complete picture of cell dynamics.

Another quantification made possible using super-resolution microscopy is the description of the cell pole form. While traditionally *C. crescentus* has been modeled as cylinder with capped hemispheres [19], the measurements in figure 30 show that the observed form varies between cells. For some cells the approximation holds well (figure 30a) while for others there is significant difference, especially at the invagination site (figure 30b). Apart from intrinsic biological variability the reasons for this difference remain unclear.

Another small feature to be considered for the description of *C. crescentus* is the cell stalk. As described in section 3.3 cells grow a stalk during replication. This stalk may persist and grow during the course of multiple cell cycles. The stalk may therefore not directly depend on the position of the cell within the cell cycle and is removed before the analysis of STORM measurements. Since SIM measurements are performed on synchronized swarmer cells, the potential contribution of the stalk to the cell length is much smaller.

6.2 Single cell identification

Observation of single cell trajectories in SIM has indicated that the developmental laws stay the same between cells while their rate constants differ (figure 31a). Their developmental trajectories overlap significantly, allowing the construction of population averages (figure 31b). The population noise therefore prohibits the reconstruction of single-cell trajectories from populations but allows the construction of population averages. The average of single cell averaged trajectories was found to be nearly identical to the direct population average.

The averages were observed to be similar for different measurements, indicating stability to biological as well as measurement noise (figure 40). The observations should therefore be easily reproducible.

For the construction of average shapes, a method for combining trajectories is required. Possible approaches and their trade-offs are described in section 4.2. In this thesis the averaging of shape parameters is used, as it leads to minimal distortions of the relevant shape features.

6.3 Pseudo-temporal classification

Pseudo-temporal classification allows to estimate the relative position of a cell within the cell cycle without requiring direct time measurements. For the STORM measurements used in this thesis such a technique is required to obtain bacterial dynamics, since the cells were fixed prior to microscopy. To assess the quality of the reconstruction the methods were performed on the SIM measurements, for which a trajectory based on measured time was available

A range of techniques described in section 3 was tested. This included generic and well known techniques like length ordering and dimensionality reduction as well as methods specifically developed for temporal reconstruction of cell cycle trajectories like *cycler*. The results were rather surprising.

The method performing best on the majority of the cell cycle was traditional length ordering. This could be due to a number of reasons. First, the used quantity of length has been observed to be the most influential quantity for pseudo-temporal classification (figure 32) and has the lowest relative measurement uncertainty of all shape parameters (figure 31a). Second, rank ordering statistics tends to be robust against outliers, locally limiting their influence to

the beginning and end of the cell cycle. The resulting average trajectory corresponds well to the population average based on measured time (figure 41).

The performance of the remaining pseudo-temporal ordering techniques turned out to be less consistent for the provided shape measurements than length ordering. A reason for this observed difference could be that these methods derive information from the measurement distribution, which may vary between populations of the same strain. In addition, figure 32 shows that other parameters are significantly less important than length, adding noise to the calculations. Since all methods apart from length ordering require at least two shape parameters the noise might out-weight the additional parameters benefit. These conclusions are also summarized in table 1.

Table 1: Comparison of pseudotemporal methods

Name	Features		Performance		Noise	Comments
	One	Multiple	End	Overall		
Length ordering	✓	✗	+	++	++	Simple method, extension to multiple parameters not straight forward
Diffusion time	✗	✓	++	+	+	Increased sensitivity to noise with few parameters
Wishbone time	✗	✓	-	-	-	Using two parameters is possible but leads to bad results in practise

Because length ordering averages of the same strain were observed to be similar for different cell populations (figure 40), this method could potentially be used for a quick screening of developmental trajectories of new strains. When applying the corrections of ergodic rate analysis no cell synchronization is necessary, allowing for quick cell dynamics estimation using fixed cell images.

7 Conclusion

Super-resolution microscopy was shown to be an important tool for the accurate description of cell division dynamics of *C. crescentus*. With diffraction-limited techniques the invagination depth can only be observed inaccurately, obscuring an important part of cell dynamics.

Of the methods tested for pseudo-temporal reconstruction of cell cycle time length ordering generally achieved good reconstruction performance for the majority of the cell cycle. The inclusion of waist width together with the diffusion maps algorithm showed an improved performance for the last fraction of the cell cycle while performing inconsistently for the remaining part. The usage of rank ordering together with the generic assumption that cells

elongate during the course of the cell cycle led to robust method applicable to a wide range of cells.

Furthermore, population trajectories generated by pseudo-temporal reconstruction have been found to be similar between different populations of the same strain, indicating usability for the description of average population dynamics.

8 Outlook

The good performance of length ordering across various bacterial strains indicate that the method could be used as technique to easily quantify the population average behaviour of new strains. In combination with the ERA correction unsynchronized bacteria can be used for microscopy, simplifying sample preparation drastically. It would therefore allow for quick screening of cell dynamics for new strains, drastically reducing cost and experimental burden for an exploratory approach to cell dynamics.

For this method to be used in a wider context, additional comparisons would have to be performed in order to ensure that the pseudo-temporal average matches the dynamics of the observed bacteria type. Evidence in this thesis indicates that this is the case of *C. crescentus*.

In addition, it has been found that the shape of the poles reassembles the classical assumptions of capped hemispheres only for some cells, while others deviate significantly from this proposed form. One possible reason for this are differences in the constriction duration of individual cells, where a smaller duration would lead to a blunter pole [2]. To investigate the exact mechanisms how this variability in pole shapes arises additional measurements would be needed.

References

- [1] Ernst Abbe. “Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung”. In: *Archiv für Mikroskopische Anatomie* 9.1 (Dec. 1873), pp. 456–468. DOI: 10.1007/bf02956176. URL: <https://doi.org/10.1007/bf02956176>.
- [2] Anna Archetti et al. Ambroise Lambert Aster Vanhecke. “Constriction rate modulation can drive bacterial cell size control and homeostasis”. 2017.
- [3] Kellie J. Archer and Ryan V. Kimes. “Empirical characterization of random forest variable importance measures”. In: *Computational Statistics & Data Analysis* 52.4 (2008), pp. 2249–2260. ISSN: 0167-9473. DOI: <https://doi.org/10.1016/j.csda.2007.08.015>. URL: <http://www.sciencedirect.com/science/article/pii/S0167947307003076>.
- [4] Mark Bates, Bo Huang, and Xiaowei Zhuang. “Super-resolution microscopy by nanoscale localization of photo-switchable fluorescent probes”. In: *Current Opinion in Chemical Biology* 12.5 (Oct. 2008), pp. 505–514. DOI: 10.1016/j.cbpa.2008.08.008. URL: <http://dx.doi.org/10.1016/j.cbpa.2008.08.008>.
- [5] Cristina Bertocchi et al. “Nanoscale Imaging by Superresolution Fluorescence Microscopy and Its Emerging Applications in Biomedical Research”. In: *Critical Reviews in Biomedical Engineering* 41.4-5 (2013), pp. 281–308. DOI: 10.1615/critrevbiomedeng.2014010448. URL: <https://doi.org/10.1615/critrevbiomedeng.2014010448>.
- [6] Ariane Briegel et al. “Location and architecture of the *Caulobacter crescentus* chemoreceptor array”. In: *Molecular Microbiology* 69.1 (July 2008), pp. 30–41. DOI: 10.1111/j.1365-2958.2008.06219.x. URL: <https://doi.org/10.1111/j.1365-2958.2008.06219.x>.
- [7] Ronald R. Coifman and Stéphane Lafon. “Diffusion maps”. In: *Applied and Computational Harmonic Analysis* 21.1 (2006), pp. 5–30. ISSN: 1063-5203. DOI: <http://dx.doi.org/10.1016/j.acha.2006.04.006>. URL: <http://www.sciencedirect.com/science/article/pii/S1063520306000546>.
- [8] Carla Coltharp and Jie Xiao. “Superresolution microscopy for microbiology”. In: *Cellular Microbiology* 14.12 (Oct. 2012), pp. 1808–1818. DOI: 10.1111/cmi.12024. URL: <https://doi.org/10.1111/cmi.12024>.
- [9] Kenneth R. Spring & Michael W. Davidson. *Introduction to Fluorescence Microscopy*. URL: <https://www.microscopyu.com/techniques/fluorescence/introduction-to-fluorescence-microscopy> (visited on 08/14/2017).
- [10] Michael W. Davidson. *Numerical Aperture*. URL: <https://www.microscopyu.com/microscopy-basics/numerical-aperture> (visited on 09/01/2017).
- [11] Kyle M. Douglass et al. “Super-resolution imaging of multiple cells by optimized flat-field epi-illumination”. In: *Nature Photonics* 10.11 (Oct. 2016), pp. 705–708. DOI: 10.1038/nphoton.2016.200. URL: <https://doi.org/10.1038/nphoton.2016.200>.
- [12] Arthur D Edelstein et al. “Advanced methods of microscope control using μ Manager software”. In: *Journal of Biological Methods* 1.2 (Nov. 2014), p. 10. DOI: 10.14440/jbm.2014.36. URL: <https://doi.org/10.14440/jbm.2014.36>.

- [13] Egelberg. *Working principle of phase contrast microscopy*. URL: https://upload.wikimedia.org/wikipedia/commons/2/23/Working_principle_of_phase_contrast_microscopy.gif (visited on 08/10/2017).
- [14] Thomas J. Fellers and Michael W. Davidson. *Introduction to Confocal Microscopy*. URL: <http://olympus.magnet.fsu.edu/primer/techniques/confocal/confocalintro.html> (visited on 08/11/2017).
- [15] Kevin Fogarty. *pgFocus*. URL: <http://big.umassmed.edu/wiki/index.php/PgFocus> (visited on 08/30/2017).
- [16] E. D. Goley, A. A. Iniesta, and L. Shapiro. “Cell cycle regulation in *Caulobacter*: location, location, location”. In: *Journal of Cell Science* 120.20 (Oct. 2007), pp. 3501–3507. DOI: 10.1242/jcs.005967. URL: <https://doi.org/10.1242%2Fjcs.005967>.
- [17] Gabriele Gut et al. “Trajectories of cell-cycle progression from fixed cell populations”. In: *Nature Methods* 12.10 (Aug. 2015), pp. 951–954. DOI: 10.1038/nmeth.3545. URL: <https://doi.org/10.1038%2Fnmeth.3545>.
- [18] Laleh Haghverdi, Florian Buettner, and Fabian J. Theis. “Diffusion maps for high-dimensional single-cell analysis of differentiation data”. In: *Bioinformatics* 31.18 (2015), p. 2989. DOI: 10.1093/bioinformatics/btv325. URL: <http://dx.doi.org/10.1093/bioinformatics/btv325>.
- [19] Leigh K. Harris and Julie A. Theriot. “Relative Rates of Surface and Volume Synthesis Set Bacterial Cell Size”. In: *Cell* 165.6 (June 2016), pp. 1479–1492. DOI: 10.1016/j.cell.2016.05.045. URL: <https://doi.org/10.1016%5C%2Fj.cell.2016.05.045>.
- [20] Jackson ImmunoResearch Inc. *Alexa Fluor 647 Technical Information*. URL: <https://www.jacksonimmuno.com/technical/products/conjugate-selection/alexa-fluor/647> (visited on 09/18/2017).
- [21] Mitchison JM. *The Biology of the Cell Cycle*. Cambridge University Press, 1971.
- [22] Stanley A. Schwartz John R. Allen Joel S. Silfies and Michael W. Davidson. *Single-Molecule Super-Resolution Imaging*. URL: <https://www.microscopyu.com/techniques/super-resolution/single-molecule-super-resolution-imaging> (visited on 09/01/2017).
- [23] Aurélie Jost and Rainer Heintzmann. “Superresolution Multidimensional Imaging with Structured Illumination Microscopy”. In: *Annual Review of Materials Research* 43.1 (July 2013), pp. 261–282. DOI: 10.1146/annurev-matsci-071312-121648. URL: <https://doi.org/10.1146/annurev-matsci-071312-121648>.
- [24] Ran Kafri et al. “Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle”. In: *Nature* 494.7438 (Feb. 2013), pp. 480–483. DOI: 10.1038/nature11897. URL: <https://doi.org/10.1038/nature11897>.
- [25] *Light Microscope*. URL: <http://symposcium.com/light-microscope/> (visited on 08/29/2017).
- [26] John R. Allen Mats G. L. Gustafsson and Michael W. Davidson. *Superresolution Structured Illumination Microscopy*. URL: <http://zeiss-campus.magnet.fsu.edu/articles/superresolution/supersim.html> (visited on 07/06/2017).

- [27] Nicolas Olivier et al. “Resolution Doubling in 3D-STORM Imaging through Improved Buffers”. In: *PLoS ONE* 8.7 (July 2013). Ed. by Markus Sauer, e69004. DOI: 10.1371/journal.pone.0069004. URL: <https://doi.org/10.1371/journal.pone.0069004>.
- [28] Martin Ovesný et al. “ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging”. In: *Bioinformatics* 30.16 (May 2014), pp. 2389–2390. DOI: 10.1093/bioinformatics/btu202. URL: <https://doi.org/10.1093/bioinformatics/btu202>.
- [29] Martin Ovesný et al. “ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging”. In: *Bioinformatics* 30.16 (May 2014), pp. 2389–2390. DOI: 10.1093/bioinformatics/btu202. URL: <https://doi.org/10.1093/bioinformatics/btu202>.
- [30] Fernando Pérez and Brian E. Granger. “IPython: a System for Interactive Scientific Computing”. In: *Computing in Science and Engineering* 9.3 (May 2007), pp. 21–29. ISSN: 1521-9615. DOI: 10.1109/MCSE.2007.53. URL: <http://ipython.org>.
- [31] Galina Reshes et al. “Cell Shape Dynamics in Escherichia coli”. In: *Biophysical Journal* 94.1 (Jan. 2008), pp. 251–264. DOI: 10.1529/biophysj.107.104398. URL: <https://doi.org/10.1529/biophysj.107.104398>.
- [32] *Resolving power (angular resolution)*. URL: <http://crab0.astr.nthu.edu.tw/~hchang/ga1/ch06-01.htm> (visited on 09/18/2017).
- [33] Rob Phillips Ron Milo. *Cell biology by the numbers*. New York, NY: Garland Science, Taylor & Francis Group, 2016. ISBN: 978-0815345374.
- [34] “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [35] Manu Setty et al. “Wishbone identifies bifurcating developmental trajectories from single-cell data”. In: *Nature Biotechnology* 34.6 (May 2016), pp. 637–645. DOI: 10.1038/nbt.3569. URL: <https://doi.org/10.1038/nbt.3569>.
- [36] Jeffrey M. Skerker and Michael T. Laub. “Cell-cycle progression and the generation of asymmetry in *Caulobacter crescentus*”. In: *Nature Reviews Microbiology* 2.4 (Apr. 2004), pp. 325–337. DOI: 10.1038/nrmicro864. URL: <https://doi.org/10.1038/nrmicro864>.
- [37] Oleksii Sliusarenko et al. “High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics”. In: *Molecular Microbiology* 80.3 (Mar. 2011), pp. 612–627. DOI: 10.1111/j.1365-2958.2011.07579.x. URL: <http://dx.doi.org/10.1111/j.1365-2958.2011.07579.x>.
- [38] Eva Wegel et al. “Imaging cellular structures in super-resolution with SIM, STED and Localisation Microscopy: A practical comparison”. In: *Scientific Reports* 6.1 (June 2016). DOI: 10.1038/srep27290. URL: <https://doi.org/10.1038/srep27290>.
- [39] R. J. Wheeler. “Analyzing the dynamics of cell cycle processes from fixed samples through ergodic principles”. In: *Molecular Biology of the Cell* 26.22 (Nov. 2015), pp. 3898–3903. DOI: 10.1091/mbc.e15-03-0151. URL: <https://doi.org/10.1091/mbc.e15-03-0151>.

- [40] A. J. M. Wollman et al. “From Animaculum to single molecules: 300 years of the light microscope”. In: *Open Biology* 5.4 (Apr. 2015), pp. 150019–150019. DOI: 10.1098/rsob.150019. URL: <https://doi.org/10.1098/rsob.150019>.
- [41] Charles S. Wright et al. “Intergenerational continuity of cell shape dynamics in *Caulobacter crescentus*”. In: *Scientific Reports* 5.1 (Mar. 2015). DOI: 10.1038/srep09155. URL: <https://doi.org/10.1038/srep09155>.

