

Diss. ETH No. 19190

COMPUTATIONAL METHODS FOR ANALYZING
AND SIMULATING INTRA-CELLULAR TRANSPORT
PROCESSES

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
JO ARNE HELMUTH
Dipl. Masch.-Ing. ETH
born on July 20th, 1979
citizen of Germany

accepted on the recommendation of
PROF. IVO F. SBALZARINI
PROF. GÁBOR SZÉKELY
PROF. GAUDENZ DANUSER
DR. MARIO SCHELHAAS

2010

ABSTRACT

Intra-cellular transport systems are a key component of eukaryotic cells. They are involved in the uptake of extra-cellular material (and pathogens) and the targeted, directed transport of such cargo inside the cell. Due to the complexity of this system and the wealth of experimental data involved in its study, computational methods for data analysis and modeling of the hypothetical system dynamics are crucial to create new knowledge.

This thesis presents a coherent set of novel computational methods and models for studying the intra-cellular transport system in mammalian cells. Three cases are considered: the intra-cellular motion of point-like organelles and cargo, the motion, shape, and dynamic shape changes of extended organelles, and spatial interactions of internalized cargo and organelles of the endocytic system.

Due to the limited resolution of live cell microscopy, the analysis of the dynamics of many small intra-cellular objects, such as vesicles and internalized viruses, is restricted to studying how their centroid position changes over time. A multitude of biophysical and biochemical processes is involved in these objects' motion, which causes the resulting trajectories to be highly patterned. A novel trajectory segmentation algorithm is presented that enables decomposing arbitrary trajectories into disjoint parts of distinct, pre-defined motion types.

For larger intra-cellular objects, such as endosomes or lysosomes, some properties of their shapes can be reconstructed from fluorescence microscopy images. The blurring inherent to fluorescence microscopy, however, obscures many details of the shape. A strong correspondence between image and object features, as required by classical image segmentation techniques, can thus not be assumed. In order to circumvent this limitation, a novel image analysis technique is introduced that combines image deconvolution and active contour segmentation. As demonstrated on benchmark data, this method is capable of accurately reconstructing the outlines of intra-cellular organelles. The new method is used in several studies of endocytic transport and sorting processes. The results indicate that endosome shape features are related to virus entry processes.

Statistical co-dependencies in the spatial distributions of two sets of objects are indicative of an interaction between them. A novel statistical framework is presented that enables testing for the presence of significant interactions and estimating their strengths and distance dependences directly from image data. The presented method thus generalizes classical object-based co-localization analysis to spatial interaction analysis. As demonstrated on locations of virus particles and computationally reconstructed outlines of endosomes, the present interaction analysis allows characterizing endocytic pathways of different virus strains.

The three aspects of analyzing endocytosis and intra-cellular trafficking presented in this thesis are connected through their relation to active, motor-dependent transport. A mechano-chemical model of active cargo transport is thus introduced. The model resolves individual motor proteins and their dynamics down to the level of single reactions, such as unbinding from and stepping along filaments of the cytoskeleton. Using exact hybrid stochastic-deterministic numerical simulations, the model is analyzed in a large-scale parameter screen. This reveals previously unknown relations between properties of the motor-cargo complex and intra-cellular transport characteristics.

ZUSAMMENFASSUNG

Intrazelluläre Transportsysteme sind eine essentielle Komponente eukaryotischer Zellen. Zum Beispiel spielen sie eine wichtige Rolle bei der Aufnahme von extrazellulärem Material (und Pathogenen) in die Zelle und dessen zielgerichteten Transport darin. Die intrazellulären Transportsysteme selbst, aber auch die Daten, die in deren experimenteller Untersuchung generiert werden, sind sehr komplex. Um aus solchen Daten neues Wissen zu schaffen, sind rechnergestützte Verfahren zur Datenanalyse sowie physikalische Modelle der hypothetischen Systemdynamik unerlässlich.

Die vorliegende Doktorarbeit befasst sich mit der Entwicklung von rechnergestützten Verfahren sowie von statistischen und physikalischen Modellen zur Untersuchung von Endozytose und intrazellulärem Transport in Säugetierzellen. Drei Teilaspekte werden dabei betrachtet: Erstens, die intrazelluläre Bewegung punktförmiger Objekte. Zweitens, die Bewegung, die Form, und die dynamische Veränderung der Form von räumlich ausgedehnten Organellen. Drittens, statistische Abhängigkeiten zwischen den räumlichen Anordnungen von internalisierten Objekten und Organellen des endozytischen Systems.

Das Auflösungsvermögen von lichtmikroskopischen Techniken, die zur Analyse von lebenden Zellen verwendet werden können, ist begrenzt. Für zahlreiche kleine intrazelluläre Objekte, zum Beispiel Vesikel oder internalisierte Viren, können deshalb lediglich die Position und deren zeitliche

Veränderung bestimmt werden. Eine Vielzahl biophysikalischer und biochemischer Prozesse ist an der Bewegung derartiger Objekte beteiligt. Dies hat zur Folge, dass die qualitativen Eigenschaften der resultierenden Bewegung starken zeitlichen Schwankungen unterliegen. In dieser Arbeit wird ein neuartiges Verfahren vorgestellt, mit dessen Hilfe Trajektorien in disjunkte Teilstücke, die hinsichtlich der qualitativen Eigenschaften der beobachteten Bewegung einheitlich sind, zerlegt werden können. Dies ermöglicht es, die den verschiedenen Bewegungen zugrunde liegenden Prozesse separat zu untersuchen.

Einige Eigenschaften der Form weniger kleiner intrazellulärer Objekte, wie zum Beispiel Endosomen und Lysosomen, können hingegen aus lichtmikroskopischen Aufnahmen lebender Zellen rekonstruiert werden. Klassische Verfahren der Bildsegmentierung versuchen Strukturen, zum Beispiel Kanten, in den Bildern zu erkennen. Anschliessend werden diese Bildstrukturen mit Objektstrukturen in Verbindung gebracht. In Bildern aus fluoreszenzmikroskopischen Verfahren werden die Objektformen jedoch stark unscharf abgebildet. Deshalb kann ein direkter Zusammenhang zwischen Bild- und Objekteigenschaften nicht unmittelbar angenommen werden. Um diese Einschränkung zu umgehen, wird ein neues Bildanalyseverfahren vorgestellt, das Dekonvolution und Segmentierung des Bildes mit Hilfe von „active contours“ kombiniert. Durch dieses Verfahren ermittelte Umrisse ausgedehnter Organellen sind weniger verfälscht als solche, die auf klassischen Ansätzen basieren. Das entwickelte Verfahren wird in mehreren Studien endozytischer Transport- und Sortierprozesse angewandt. Die Resultate suggerieren, dass die Formeigenschaften der untersuchten Organellen mit biophysikalischen und biochemischen Prozessen korrelieren.

Statistische Abhängigkeiten in der räumlichen Anordnung von zwei Gruppen diskreter Objekte können ein Hinweis darauf sein, dass die Objekte miteinander interagieren. Es wird ein neues statistisches Verfahren vorgestellt, das es ermöglicht derartige Interaktionen direkt aus Bildern der Objekte zu schätzen bzw. zu testen, ob eine signifikante Interaktion vorhanden ist. Das neue Verfahren kann als eine Verallgemeinerung der klassischen objektbasierten Kolokalisationsanalyse interpretiert werden. Es ermöglicht es, die endozytischen Wege verschiedener Mutanten humaner Adenoviren anhand der aus Bildern rekonstruierten Positionen von Endosomen und Viruspartikeln zu charakterisieren.

In dieser Arbeit werden drei Teilaspekte einer ganzheitlichen Analyse von Endozytose und intrazellulärem Transport vorgestellt. Jeder dieser drei Aspekte hat einen engen funktionalen Zusammenhang mit aktivem intrazellulärem Transport durch Motorproteine. Deswegen wird ferner ein mechanisch-chemisches Modell des aktiven intrazellulären Transports vorgestellt. Das Modell beschreibt die Dynamik von Motorproteinen auf der Ebene einzelner chemischer Reaktionen. Anhand exakter stochastisch-deterministischer numerischer Simulationen wird das Modell für eine Vielzahl verschiedener Parameter analysiert. Zuvor unbekannte Zusammenhänge zwischen Eigenschaften des Motor-Fracht Komplexes und der Transportcharakteristiken werden aufgedeckt.

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Ivo Sbalzarini for his great support. Ivo was always encouraging, generous, fair, open to new ideas, flexible, and extremely responsive, all of which are qualities that can't be taken for granted. I feel lucky that Ivo triggered my interest in computational science and its application in cell biology. Thanks to Ivo I could get to know many interesting people inside academia. I very much enjoyed the scientific exchange with our friends and collaborators in Paris and Split.

I am grateful that I had the opportunity to participate in the buildup of the MOSAIC group. Soon after the beginning of my PhD in summer 2006, when I was the only student in Ivo's group, Christian and Birte joined. Already then, we were a fairly heterogeneous crowd, but the heterogeneity grew with every new member that joined. I am thankful that we, the PhD students, could participate in selecting new group members – this was a great privilege. The group's heterogeneity had a tremendous effect on the group culture, which I perceived as positive, amicable and professional. I believe that, after a while, this heterogeneity had a positive effect on our creativity and productivity. It was great fun and a good experience being a part of a MOSAIC of exclusively exceptional people.

Special thanks go to my mentors, office mates, and collaborators within and outside the group. I very much appreciated Janick's positive attitude

ACKNOWLEDGMENTS

and pragmatic support as well as Greg's detailed, critical, and constructive advices. Without his knowledge and ideas, our joint project on interaction analysis would not have been possible. Sylvain deserves special mentioning for structuring my sometimes convoluted lines of thought and his contribution to development of the transport model. Rajesh's endurance and patience were a positive example for me. Christian's enthusiasm, which is paired with a great deal of thoughtfulness, helped me stay motivated in less enjoyable times of my PhD. Birte always provided thorough support no matter how busy she was herself.

With Christoph Burckhardt I had a strong ally among the cell biologists. Our collaboration, which lead to two joint publications, was very close and interactive. Christoph's open-mindedness made this possible. Special thanks go also to Prof. Urs Greber, Christoph's former PhD advisor. Mirko Birbaumer made me become interested in modeling intra-cellular transport processes. Credit goes to Mattia Gazzola for pointing out that modeling forces involved in active transport is challenging. I furthermore would like to acknowledge the contributions of the students I supervised during my time at ETH. Without the ideas of Alain Weber, Benedikt Baumgartner, and Markus Sutter my thesis would not have been the same.

I am very grateful to my co-examiners, Prof. Gábor Székely, Prof. Gaudenz Danuser, and Dr. Mario Schelhaas for reviewing my thesis. I feel honored by their interest in my work.

CONTENTS

ABSTRACT	I
ZUSAMMENFASSUNG	III
ACKNOWLEDGMENTS	VII
INTRODUCTION	XIII
1 QUANTIFYING THE DYNAMICS OF POINT-LIKE INTRA-CELLULAR OBJECTS	1
1.1 RECORDING AND RECONSTRUCTING TRAJECTORIES	3
1.2 GLOBAL TRAJECTORY STATISTICS	4
1.2.1 Local Trajectory Statistics	6
1.3 TRAJECTORY SEGMENTATION	6
1.3.1 Definitions and Problem Statement	8
1.3.2 Algorithm	9

CONTENTS

1.3.3	Adaptation for Segmenting Virus Trajectory Data . . .	20
1.3.4	Evaluation	27
1.3.5	Application	32
2	QUANTIFYING THE SHAPE AND DYNAMICS OF EX- TENDED INTRA-CELLULAR OBJECTS	41
2.1	MICROSCOPY, DECONVOLUTION, POINT SPREAD FUNC- TIONS, ETC.	43
2.1.1	Fluorescence in Cell Biology	43
2.1.2	Basics of Image Formation	44
2.1.3	Resolution and Blurr	49
2.1.4	Digital Images and Noise	50
2.1.5	Deconvolution	52
2.1.6	Empirical Characterization of Microscopes	55
2.2	SEGMENTATION OF INTRA-CELLULAR STRUCTURES	57
2.2.1	Definitions and Problem Statement	60
2.2.2	Algorithm	60
2.2.3	Evaluation	74
2.2.4	Applications	81
3	QUANTIFYING INTERACTIONS BETWEEN INTRA-CELLULAR OBJECTS	99
3.1	INTENSITY CORRELATIONS	101
3.2	OBJECT-BASED INTERACTION ANALYSIS	103

3.2.1	Classical Co-localization Measures	104
3.2.2	Theoretical Distance Distribution	109
3.2.3	Statistical Toolbox I: Hypothesis Testing	114
3.2.4	Statistical Toolbox II: Estimation of Potentials	123
3.2.5	Application of the Method	131
4	MODELING THE MULTI-SCALE TRANSPORT DYNAMICS OF INTRA-CELLULAR OBJECTS	147
4.1	A TUG-OF-WAR TRANSPORT MODEL	150
4.1.1	Specification of the Model	152
4.1.2	Simulating the Model	157
4.2	STUDYING THE MODEL	163
4.2.1	Parameters	165
4.2.2	Auto-Correlation of the Cargo Velocity	166
4.2.3	Multi-Scale Dynamics	168
4.2.4	Persistence of Uni-Directional Motion	171
4.2.5	Velocity Distributions	175
4.3	RELATIONSHIPS TO OTHER MODELS	182
5	CONCLUSIONS AND OUTLOOK	185
5.1	TRAJECTORY SEGMENTATION	186
5.2	DECONVOLVING ACTIVE CONTOUR SEGMENTATION OF INTRA-CELLULAR ORGANELLES	188
5.3	INTERACTION ANALYSIS FOR INTRA-CELLULAR OBJECTS	190

CONTENTS

5.4 ACTIVE TRANSPORT MODEL	192
5.5 SIGNIFICANCE OF THE PRESENT WORK	194
BIBLIOGRAPHY	199
LIST OF OWN PUBLICATIONS	215

INTRODUCTION

Cells rely on highly structured and complex mechanisms for processing signals, interacting with the environment, performing metabolism, growing, dividing, etc. These mechanisms often depend on a tight coupling between multiple biochemical and biophysical processes. For endocytosis and intracellular trafficking, for example, it becomes increasingly clear that a tight coupling between signaling, physical processes such as diffusion or active transport, and the cell output exists.

As cell biology is undergoing a transition from a descriptive to a quantitative and predictive science, a system-wide view is adopted in order to allow gaining new insight beyond a purely qualitative description of the individual parts that constitute a cell. In this approach, identification and quantification of the biophysical and biochemical interactions between the parts plays a major role. Precise quantification of the properties of the individual parts, however, remains a fundamentally important prerequisite. Coping with the overwhelming complexity of cellular systems and their dependence on physical processes necessitates approaches to data analysis and modeling that integrate biology, physics, chemistry, and computational science.

Endocytosis is the process by which the cell membrane forms invaginations to engulf and subsequently internalize material. For the endocytic system, many molecular insights have been gained [117, 137]. Its physical

INTRODUCTION

principles of operation and integration in the overall cellular machinery are, however, largely unexplored [26]. In recent years, it has become clear that endocytic processes and signaling are interdependent and mutually regulating [114, 105, 128]. Since endocytosis removes signaling molecules from the cell surface, it can attenuate signaling pathway responses. Conversely, signaling endosomes that expose active signaling molecules to the cytosol can be used to increase signaling output. By actively moving such endosomes closer to the signal target, the diffusion limitation of signaling pathway response can be overcome.

In millions of years of co-evolution with their hosts, viruses have become experts in taking advantage of the host's endocytic system. Viruses can hijack the endocytic transport machinery, which allows them to efficiently deliver their genome and accessory proteins into the cell's interior for subsequent replication [97, 114]. Different endocytic pathways lead to different organelles, such as caveosomes, early and late endosomes, lysosomes, the *trans*-Golgi network, or the endoplasmic reticulum [114]. Co-localization or other statistical co-dependencies between viruses and endocytic organelles may therefore provide hints for identifying virus entry routes. Observing a virus infection program from internalization to replication, however, also allows probing the functioning of the endocytic system itself [114, 113].

Key components of a certain cellular function are typically identified in perturbation studies. Small interfering RNA (siRNA) studies and related approaches, for example, allow systematically perturbing groups of genes. In a recent study of the endocytic system, the siRNA technique has been applied to correlating the targeted gene with multiple phenomenological parameters of the endocytic system, such as number, size, and position of endosomes [26]. A key factor in such perturbation studies is that the *measured* parameters are informative, that is, they have to yield high statistical power for detecting potential effects of gene interference. Mechanistic insight, however, further requires that the parameters can be interpreted within models of biophysical or biochemical processes.

This thesis is concerned with developing statistical and computational methods that help *design* (in the sense of experimental design), *measure* (in the sense of data analysis), and *reproduce* (in the sense of modeling

and simulation) parameters that correlate with biochemical and biophysical processes involved in endocytosis and intra-cellular trafficking. It is structured as follows:

CHAPTER 1: QUANTIFYING THE DYNAMICS OF POINT-LIKE INTRA-CELLULAR OBJECTS

The first chapter is concerned with the identification and characterization of different types of motion of small, point-like objects such as viruses. Viruses interact with their host cells in numerous ways, yielding highly heterogeneous motion trajectories. Some of the motion patterns present in virus trajectories can be attributed to specific virus–cell interactions [16].

A novel algorithm for segmenting pre-defined motion patterns from heterogeneous trajectories for subsequent analysis is presented in Section 1.3. The algorithm is based on support vector classification of trajectory features. This approach allows training the algorithm to detect pre-defined motion types without having to explicitly define the classification rules.

The algorithm’s performance is demonstrated on synthetic data and by applying it to segmenting four distinct motion types in trajectories of human Adenovirus particles in several different live cells.

CHAPTER 2: QUANTIFYING THE SHAPE AND DYNAMICS OF EXTENDED INTRA-CELLULAR OBJECTS

Chapter 2 is concerned with determining the dynamic shapes of extended objects such as endosomes. A system level perturbation study of the endocytosis based on siRNA gene interference has been recently performed [26]. Such studies crucially depend on robust, accurate, and unbiased estimation of parameters that characterize perturbed cells on the level of single organelles.

In Section 2.2, an image analysis algorithm is introduced that reconstructs the precise shapes of endosomes (or any other compact sub-cellular struc-

ture) from fluorescence microscopy images. On the length scale of the typical size of endosomes, such images severely suffer from the limited resolution of standard fluorescence microscopy. The presented algorithm is based on modeling the image formation process in fluorescence microscopy. This allows determining unbiased endosome outlines with sub-pixel accuracy.

Based on these outlines, endosome shape features are computed that allow, for example, detecting fusion events. The same endosome shape features are also shown to be correlated with the presence of Adenovirus particles that entered the cell via a clathrin-dependent endocytosis prior to imaging.

CHAPTER 3: QUANTIFYING INTERACTIONS BETWEEN INTRA-CELLULAR OBJECTS

A frequent task in cell biology is to identify spatial correlations, or *interactions*, between intra-cellular objects such as molecules, organelles, or internalized pathogens. Endocytosis of viruses, for example, is a complex process that depends on membrane receptors, regulatory molecules, signaling, specific organelles and the chemical cues inside them. The entry pathways of different viruses are typically classified according to the cellular machinery involved. The membrane receptor the virus binds to, its initial sorting compartment, or where the viral genome penetrates the membrane of intra-cellular organelles are fingerprints of the pathway taken. Evidence that a virus and an organelle *co-localize* in space and time may thus help identify its endocytic pathway.

A novel data analysis method that provides a generalization of co-localization analysis is introduced in Chapter 3. The method allows constructing interaction measures from localization patterns of intra-cellular objects. The core of the method is a statistical model for the distribution of inter-object distances. This model is an adaptation of models frequently used in the theory of spatial point patterns of interacting objects.

As demonstrated on synthetic data, the novel interaction measures have significantly higher statistical detection power than commonly used co-

localization scores. The statistical nature of the underlying model moreover enables using a wealth of standard statistical tools, such as non-parametric hypothesis tests, maximum-likelihood inference of parametric or non-parametric interaction models, model selection, etc. Using the virus–endosome data presented in Chapter 2, the novel measures are used to characterize endocytic pathways of different virus strains.

CHAPTER 4: MODELING THE MULTI-SCALE TRANSPORT DYNAMICS OF INTRA-CELLULAR OBJECTS

The heterogeneity of intra-cellular motion trajectories, virus trafficking through different pathways, and the spatial distribution and dynamic shape changes of involved objects are closely related to active transport mechanisms. Motor proteins move cargo along filaments of the cytoskeleton, over distances comparable to the cell size [146]. During active transport, motor forces can cause deformations of non-rigid cargo. This has been observed for endosomes [133], which consist of deformable lipid membranes enclosing internalized cargo. Motor forces may also play an important role in initiating fusion [42] or fission of organelles. Such topological changes of endocytic organelles are essential mechanisms for cargo sorting.

The forth chapter of this thesis is concerned with modeling and simulating active intra-cellular transport of various cargos. Using single motors as elementary building blocks, a novel tug–of–war model for intra-cellular transport along microtubules is developed. In this model, the forces acting on motors and cargo are explicitly resolved, since they are important in the stochastic mechano-chemical dynamics of collective motor motion. The model is efficiently simulated using a variant of the stochastic simulation algorithm for hybrid stochastic-deterministic chemical kinetics [49, 50].

As shown in a large-scale parameter study, experimentally observable transport properties are highly sensitive to the elasticity of the motor–cargo links and the drag force exerted on the cargo by the surrounding fluid. The present model can also be extended to account for deformations of the cargo or cargo–cargo interactions, such as traffic jams on filaments or force-induced fusion and fission.

QUANTIFYING THE DYNAMICS OF POINT-LIKE
INTRA-CELLULAR OBJECTS

This chapter describes the analysis of trajectories of intra-cellular objects.

Complex trajectories are universal hallmarks of cellular and organismal life. Trajectories of intra-cellular organelles – or in general intra-cellular objects – comprise transient patterns of diffusive motion, stalling, confinement, or directed motion. The trajectories themselves are of little value for the biological sciences. Trajectories and the patterns therein, however, result from chemical, physical, or complex regulated biological interactions between moving objects and their environment. Therefore, the acquisition and analysis of trajectories of intra-cellular objects may provide insight into the mechanisms underlying their motion. This approach has the beneficial property that it is largely non-invasive and that it can be used in *in vivo* experiments.

In cell biology, acquisition of trajectories is almost exclusively based on time-lapse microscopy, and trajectories have to be reconstructed from the resulting image data [124, 67, 125]. This problem is called *feature point*

tracking. First, the object under investigation has to be detected in the individual frames and its precise location needs to be determined. If multiple objects are present at the same time, an association across frames between the detected objects needs to be found. Originally, this process was carried out by hand in a tedious, non-reproducible, subjective, and error-prone way. The development of efficient feature point tracking algorithms that run on powerful computers automatized and objectified this process, which allowed to produce large numbers of trajectories. Furthermore, the tracking algorithm can be characterized a priori, which provides estimates for the localization and linking accuracy.

Inside cells, forces are typically of short range and interactions are limited to the immediate neighborhood of objects. Objects frequently exchange momentum with molecules in the cytoplasm through collisions. Due to the high viscosity of the cytoplasm and the low mass and velocity of the objects inertial effects can be safely neglected. For these and other reasons, the motion of intra-cellular objects is highly stochastic in nature. Formally, the motion of intra-cellular objects can be described by a *stochastic process* generating random object positions $\mathbf{x}(t)$. The characterization of such trajectories is therefore limited to determining *statistical properties*, such as the *probability density* $p(\mathbf{x}, t + \delta t | \mathbf{x}_0, t)$ for moving from \mathbf{x}_0 at time t to \mathbf{x} at a later time $t + \delta t$. Models of such stochastic process are often phenomenological, but they may as well be based on mechanistic models for the object's motion. The stochastic processes can be stationary, which motivates the use of global trajectory statistics (reviewed in Section 1.2), such as the diffusion constant. The process can also be parameterized by time, in which case local trajectory statistics (reviewed in Section 1.2.1) may reveal more detail. Finally, the trajectory may be composed of distinct segments of qualitatively different motion patterns that are separated by sharp transitions. Global and local trajectory statistics may be misleading in this case, since they are blind for such transitions. In order to avoid mixing of motion types in local or global statistics, trajectories have to be decomposed into segments containing only one single pre-defined motion pattern. In Section 1.3 I devise, test, and demonstrate a novel algorithm that solves this *trajectory segmentation* problem.

1.1 RECORDING AND RECONSTRUCTING TRAJECTORIES

Specific fluorescence labeling of intra-cellular objects allows observing their motion by means of time-lapse fluorescence microscopy. Feature point tracking is concerned with reconstructing the trajectories of moving objects from a sequence of digital images. Two main tasks are typically solved independently: First, objects have to be *detected* in the images. In case multiple objects are present in the images, the individual object locations have to be linked in time, which amounts to solving a *linear assignment problem* [125, 67]. In *particle filtering* [4] approaches, the borders between detection and linking are blurred, since a motion model (i.e. something related to linking) is used to facilitate the detection. Challenges in tracking include high object densities, temporary disappearance of objects, high velocities, erratic motion, etc. Tracking is easy, for example, when objects are well-separated, move according to a highly deterministic model, or can be distinguished based on their appearance (see Figure 1.1).

Computational approaches to feature point tracking can be loosely classified into methods that include prior knowledge about the motion of the objects, and those that don't. While including prior knowledge may help resolve some of the problems mentioned above, it limits the applicability of the tracking algorithm to specific classes of problems.

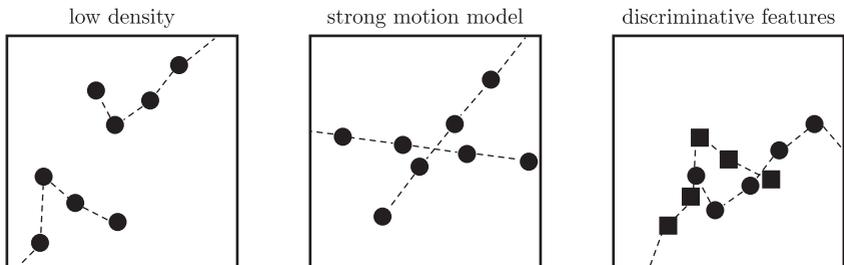


Figure 1.1: Situations that facilitate feature point tracking.

In this chapter, the algorithm in reference [125] was used to reconstruct trajectories of moving virus particles from fluorescence microscopy images.

It separately solves the tasks of detecting object locations and linking them into trajectories.

The algorithm assumes that images of single objects are isolated peaks of finite extension. First, tentative objects locations are identified as local intensity maxima by means of a gray-scale morphological operation. The object locations are then refined to sub-pixel accuracy by intensity centroid estimation. Objects can optionally be classified into correct and spurious detections based on their central intensity moments.

Each image is thus reduced to a set of point locations. Trajectories of individual objects are then found by linking these detections over time. Each link is associated a *cost* that is based on the differences in location and intensity moments. Appearing and disappearing particles can be linked to dummy particles with a fixed cost. The linking algorithm finds an overall optimal association between the object locations, that is, the trajectories are constructed such that the total cost for all trajectories is minimal.

1.2 GLOBAL TRAJECTORY STATISTICS

Many important properties of trajectories, such as the end-to-end distance, are global quantities. Furthermore, trajectories may result from a *stationary* stochastic process, which is characterized by a *transition density* for the location of the object. Such a density is often fully defined by a set of parameters. For normal isotropic diffusion, for example, this would be the *diffusion constant* D . Such parameters can be estimated from parts of the trajectory. Using all available information (i.e. the whole trajectory), however, typically reduces the variance of the estimator. End-to-end distances, diffusion constants, and alike are therefore considered *global trajectory statistics*.

Quite generally, global trajectory statistics can be defined as averages of some function of the object's location \mathbf{x} . For example, the diffusion constant D in d dimensions can be related to the *mean squared displacement*

(MSD) $\mu_2(\delta t)$ within time δt through

$$D = \frac{\langle \|\mathbf{x}(t + \delta t) - \mathbf{x}(t)\|^2 \rangle_t}{2d\delta t} = \frac{\mu_2(\delta t)}{2d\delta t}, \quad (1.1)$$

which, assuming that the process is *ergodic*, follows from the Gaussian transition density for isotropic diffusion (see Section 1.3.4.1) by integration. In the above equation $\langle \cdot \rangle_t$ denotes an average over all possible times t .

For a normal diffusion process, $\mu_2(\delta t)$ grows linearly with δt , i.e. $\mu_2 \propto \delta t^\gamma$ with $\gamma = 1$. *Anomalous diffusion* is characterized by a different growth law for $\mu_2(\delta t)$, where $\gamma \neq 1$. As a generalization of Equation 1.1, one can define finite sample estimators for the p -th displacement moment:

$$\mu_p(\delta t) = \langle \|\mathbf{x}(t + \delta t) - \mathbf{x}(t)\|^p \rangle_t. \quad (1.2)$$

As shown for the second moment, each moment obeys a scaling law [40]:

$$\mu_p(\delta t) \propto \delta t^{\gamma(p)}. \quad (1.3)$$

This allows defining the *moment scaling spectrum* (MSS) of a trajectory as the function

$$p \rightarrow \gamma(p). \quad (1.4)$$

The shape of the MSS allows inferring important properties of the diffusion process. *Strongly self-similar* diffusions have a linear MSS, and the slope β of that linear function allows further classification of the diffusion process [40]. For normal diffusion, $\beta = 1/2$. Departure from this value allows detecting sub-diffusion ($\beta < 1/2$), super-diffusion ($1/2 < \beta < 1$), and ballistic motion ($\beta = 1$). Non-linear functions $\gamma(p)$ are indicative of *weakly self similar* diffusion processes.

A complementary approach to whole-trajectory characterization was recently suggested by Wieser and co-workers [153]. They use a statistical test to compare the distribution of experimentally observed MSDs for a given time lag to the distribution obtained from a Monte-Carlo simulation of a hypothetical diffusion process.

1.2.1 LOCAL TRAJECTORY STATISTICS

Local trajectory statistics are a straightforward extension of global trajectory statistics. The idea is to capture non-stationarity of the stochastic process by computing statistics within parts of the trajectory. For computation of the statistics, the normal formulae for global statistics can be used. In a typical scenario, statistics are computed within a *moving window* that is systematically shifted over the trajectory. However, care must be taken to not over-estimate the significance of fluctuation of any statistic, since the statistics within moving windows will exhibit higher variance than global counterparts due to the reduced sample size. Correlations between the values of overlapping windows have to be considered as well, if one aims at detecting non-stationarity.

Non-stationarity of a trajectory-generating stochastic process may be detected based on local trajectory statistics computed in moving windows. Sudden changes of the statistic between neighboring windows can be indicative of a motion driven by quantitatively or qualitatively different processes. However, the *statistical detection power* of this approach is limited by the tradeoff between temporal resolution and statistical uncertainty. Furthermore, specific statistics are often only sensitive to certain deviations from “normal” behavior in the trajectories, but are completely blind to others. This has motivated to use a collection of statistics (a *bag of features*) that are specifically designed to be sensitive to different characteristics of the trajectory. Changes in the motion dynamics may become apparent in the high-dimensional feature space spanned by the different local trajectory statistics [65].

1.3 TRAJECTORY SEGMENTATION

Many intra-cellular objects exhibit highly non-stationary behavior with drastic qualitative changes of the dynamics: Free as well as confined diffusion, anomalous diffusion, diffusion with overlaid drift, fast directed transport, and stalling are some of the components that frequently alternate. In such cases it is desirable to not only *visualize* changes, but to *decompose* the observed trajectories into segments, where each segment contains

only one single pre-defined motion pattern. This allows to characterize resulting segments independently (*How fast is directed motion?*), to count their frequencies (*How often does directed motion occur?*), or to find spatial and temporal correlations of their occurrence (*Where in the cell does directed motion occur? Is directed motion frequently followed by stalling?*). For rare patterns, the segmentation can be seen as a way to remove the “background noise” from the “signal” of interest (*Is there any directed motion in the otherwise diffusive motion?*). Ideally, one is able to define and detect motion patterns that have a one-to-one correspondence with a (bio)-physical mechanism of specific interest. Fast uni-directional (or *saltatory*) motion, for example, is often attributed to molecular motors that processively move cargo along filaments of the cytoskeleton. Segmenting patterns of fast directed motion thus allows estimating physical properties of the motor-cargo complexes [44], such as average velocity, traveling distance, etc.

Trajectory segmentation amounts to solving a pattern detection or classification problem: *Given a collection of features in multiple time windows, can one conclude that they result from a specific stochastic process? To which of the known patterns do the observations most likely correspond?* It might be possible to phrase these problems in terms of model selection or statistical classification theory. However, this requires statistical models for the data, which, in case of non-trivial trajectory features, might not be readily available. Given a set of trajectories with known segmentation, one can, however, design and train a heuristic classifier. In this context, the *class* of motion is a stochastic process, that yields a specific motion pattern. In the first step of the construction of a classifier, collections of trajectory features need to be computed in parts of the trajectory within windows around different points in time. These *feature vectors* are then labeled with the type of pattern – i.e., the class of motion – that is exclusively present in the segment in which they were computed. Trained with such labeled data, the classifier might then also correctly decide on the class membership of a previously unseen feature vector. The quality of this classifier crucially depends on the choice of features, the size of the window in which they are computed, and the amount and diversity of training data. In Section 1.3.2, a general description of the present trajectory segmentation algorithm is given. Following the above-mentioned design guidelines, the algorithm is then adapted to segmenting Adenovirus

trajectories (Section 1.3.3). Sections 1.3.4 and 1.3.5 report results of an evaluation on synthetic data and an application to trajectories of Adenovirus particles moving on the plasma membrane or inside human cells.

1.3.1 DEFINITIONS AND PROBLEM STATEMENT

A *trajectory* is defined as a time series of n *positions* \mathbf{x}_i of the tracked object, sampled at discrete time intervals dt :

$$T = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \quad \text{with } \mathbf{x} \in \mathbb{R}^d. \quad (1.5)$$

From these positions one can compute the trajectory *steps* $\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$, with $i \in [1, n-1]$, leading to an equivalent representation of the trajectory as an ordered sequence of steps:

$$T' = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{n-1}\} \quad \text{with } \mathbf{s} \in \mathbb{R}^d. \quad (1.6)$$

Features p are defined as the scalar result of applying a *dimensionality reduction operator* to a sequence of l_P steps, thus:

$$\Psi : \mathbb{R}^{d \times l_P} \rightarrow \mathbb{R}, \quad p = \Psi(\{\mathbf{s}_i, \mathbf{s}_{i+1}, \dots, \mathbf{s}_{i+l_P-1}\}). \quad (1.7)$$

A vector of m such features p^i , resulting from application of multiple operators Ψ^i , is referred to as a *feature vector* $\mathbf{f} = (p^1, \dots, p^m)^T$. The different *classes* c of motion within a trajectory are assumed to be identified by characteristic feature vectors, or, more generally, by a characteristic probability density distribution in feature vector space. If the feature vectors are sensitive to position noise in the trajectories, the point positions can be *filtered* before computing the features: $\tilde{T} = \Phi(T)$. The *trajectory filter function* Φ hereby describes a unique and explicit mapping from the n point positions in T to \tilde{n} positions in \tilde{T} . Finally, a *trajectory part* P is defined as a subset of T containing $l_P + 1$ subsequent positions \mathbf{x} (and hence l_P steps \mathbf{s}) between two discrete time points. Different trajectory parts may have different lengths. The process of *trajectory segmentation* recognizes pre-defined motion patterns in the trajectory T , and subdivides T into disjoint trajectory parts P , each containing only one class of motion c . Such trajectory parts are termed *trajectory segments* S .

1.3.2 ALGORITHM

As stated in Section 1.3, the fundamental idea behind the present trajectory segmentation algorithm is to perform a classification of trajectory parts P into pre-defined classes c of motion. In the present context, the classifier is considered a “black box”. This means that the inner workings of it are not important, as long as the relation between input (a feature vector) and output (a motion class label) meets certain requirements.

A large variety of classifiers can be distinguished that are based on different principles, for instance *neural networks* [12], *support vector machines* [127], *Fisher’s linear discriminant* [37], *nearest neighbor classification* – to just name a few. All of these pursue similar objectives and face similar problems. Therefore, I want to briefly review some basic concept of classification, introduce the basics of support vector machines, and then move on to the specification of the actual trajectory segmentation algorithm.

1.3.2.1 Binary Classification

A classifier assigns a class label to some input. *Binary classification* is a special case, in which only two classes exist that the classifier can choose from. The decision is based on some characteristics of the inputs, provided in the form of a feature vector. Consider the classification of dinosaurs into the classes “herbivore” and “carnivore” as an illustrative example. A vector of features could for instance consist of measures of height and weight. Considering the great variety of sizes of both carnivores and herbivores, this is not adequate to achieve reasonable classification of any dinosaur. This exemplifies the importance of using more informative features, such as “scariness of teeth” or “sharpness of claws”. Facing a *Spinosaurus* (having scary teeth *and* sharp claws!), one would correctly conclude that it is a good idea to to back up. However, there are also plenty of herbivores that have reasonably sharp claws or even a few scary teeth. It is therefore necessary to *learn a decision boundary* using a *training set* of features of dinosaurs with known preferred nutrition. This decision boundary separates the feature space into two or more regions, each corresponding to either carnivores or herbivores (Figure 1.2). Depending on the type of

classifier used, the decision boundary could simply be a line, with potentially some of the dinosaurs ending up on the wrong side. This situation is called *under-fitting*. It is a risky business, because one knows for certain that some dinosaurs will be mis-classified in the future. Allowing for more complex, non-linear decision boundaries, one could possibly separate the training data perfectly. This is, however, another risky situation called *over-fitting*. The problem is that the classifier has learned too many local details, at the expense of identifying less of the significant global trends. Fundamentally, the objective is to *generalize* well, that is, to classify future, unknown data as well as possible (solid line in Figure 1.2).

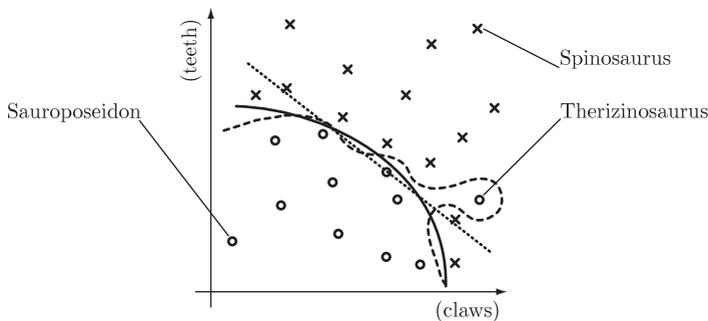


Figure 1.2: Classification of dinosaurs into carnivores (crosses) and herbivores (circles) using physiological features. The lines depict decision boundaries; dotted: under-fitting, dashed: over-fitting, solid: reasonable generalization.

1.3.2.2 Multi-Class Classification

Multi-class classification is concerned with grouping objects into one of more than two classes. Consider again the classification of dinosaurs. This time, however, the order of “Saurischia” shall be further divided into the classes “Heererasauria”, “Theropoda”, and “Sauropodomorpha”.

As opposed to binary classification, the methodology and formalism of multi-class classification is much less developed. A common approach is

to reduce the problem to a set of binary classification problems, for example to a specific one-versus-all problem per class. The prerequisite for this is that for each class a criterion exists that allows unambiguous positive identification. Among the classes mentioned, the Sauropodomorpha can, for example, be identified by their long neck. Positive identification of Heererasauria and Theropoda, however, is less trivial since they share many characteristics. If, nevertheless, criteria for the identification Heererasauria and Theropoda can be found, the binary classifications can be performed in any order, since, by definition, no object can be assigned to more than one class.

A classification criterion may not easily be found when one class occupies a small portion of the part of feature space that is normally attributed to another class. In this particular case, the binary classifications can be performed in a hierarchical way: After the first positive identification, the object it is no longer considered in any of the following binary classifications. It might happen, though, that none of the classes are clearly identified, leaving objects unlabeled. A trivial fix is to assign the label of the most general class to all unidentified objects, or to group them in an auxiliary class.

1.3.2.3 Support-Vector Machines for Trajectory Segmentation

Support-vector machines [127] (SVMs) are powerful classifiers that have been successfully applied in countless applications. SVMs separate classes by a hyperplane in a high-dimensional space. This hyperplane is constructed such that it maximizes the distance to the closest points on each side of it (see Figure 1.3). Such a maximum *functional margin* should be beneficial for generalization.

Non-linear decision boundaries are realized by transforming the features into a high-dimensional space using a non-linear kernel function and solving a linear classification problem in this space. This is equivalent to having a non-linear decision boundary in the original feature space (see Figure 1.4).

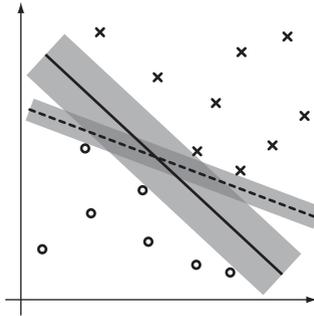


Figure 1.3: Functional margin of linear classifiers: The dashed decision boundary separates the two classes perfectly, the solid boundary, however, has a larger functional margin (shaded areas).

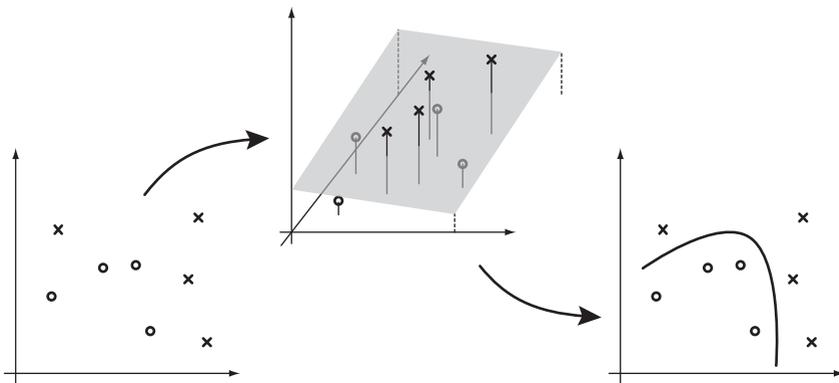


Figure 1.4: Support vector machines: After a non-linear mapping into a high-dimensional feature space, linear classifiers can realize non-linear decision boundaries in the original feature space.

In many situations the classes are not separable, or only with highly convoluted decision boundaries, which are of little practical use due to their reduced generalization. The maximum margin concept can then be relaxed to *soft margins* by introducing a penalty for misclassification. The penalty is typically proportional to the distance from the boundary. For a given (non-linear) kernel function, the boundary that minimizes the penalty for all misclassified data, while maximizing the margin for correctly classified data, is considered optimal. In order to combine these two objectives into a single *objective function* a scalar weighting parameter C for the penalty is introduced. The resulting constrained optimization problem is usually solved *quadratic programming*, a non-trivial, but very mature technique. In the context of support vector machines, “training” hence refers to the process of solving the quadratic programming problem for a given set of feature vectors with known classes [127].

Here, I use a C -SVM with a radial basis function kernel for the classification of trajectory features. The kernel has a parameter σ that, together with the penalty weight C , controls the complexity of the decision boundary and therefore the potential to generalize well. In practice the parameters $\{C, \sigma\}$ need to be set a priori or selected during the training process.

1.3.2.4 Cross-Validation: Assessing Generalization

The purpose of *cross-validation* is to quantify the ability of a classifier to generalize after training. This is achieved by using only a subset of the available training data and keeping the rest for validation. On this *validation set*, the classification performance is assessed immediately after training. Motivated by detection problems in, e.g., medical diagnosis, the two classes are called “positive” and “negative”. The classification performance is then expressed in terms of the ratio of true positives to all positives (the *sensitivity*), and the ratio of true negatives to all negatives (the *specificity*). Resolving the performance for both classes separately has the advantage that biases toward one class (for instance if the training set comprises different numbers of samples of the two classes) can be detected.

In k -fold cross-validation the original training data is split into k equally large sets. Each of the k sets is used once as validation set, when the classifier has been trained on the remaining $k - 1$ sets. In case of the SVMs used for the trajectory segmentation algorithm, this procedure is repeated with different parameters $\{C, \sigma\}$ (see Section 1.3.2.3) and the combination with the best average performance (e.g., highest average sensitivity and specificity) is selected.

1.3.2.5 Specification of the Algorithm

The trajectory segmentation algorithm consists of four steps:

1. Trajectory preprocessing
2. Feature extraction
3. Feature classification
4. Classifier output filtering and label assignment

Step 1: Trajectory preprocessing. The purpose of trajectory preprocessing is to remove position noise from the data by means of a trajectory filter function Φ . Furthermore it allows setting a time scale for the feature extraction a priori, which may be beneficial for patterns corresponding to slow processes.

In the absence of detailed knowledge about the position noise, the simplest choice for Φ is a box filter with width w :

$$\Phi : T \rightarrow \tilde{T}, \quad \tilde{\mathbf{x}}_i = \frac{1}{w} \sum_{j=(i-1)w+1}^{iw} \mathbf{x}_j. \quad (1.8)$$

Trajectory preprocessing is an optional step. There is no restriction regarding the use of different filters for different patterns. In fact, in case of trajectories containing multiple patterns with different typical time scales, feature extraction may be performed on separate filtered trajectories for each pattern.

Step 2: Feature extraction. Trajectory parts are transformed to low-dimensional feature vectors by application of reduction operators Ψ (Equation 1.7). The operators are applied to all possible consecutive trajectory parts of length l_P , that is, features are computed from the content of a window of length l_P moved over the (filtered) trajectory step by step (see Figure 1.5).

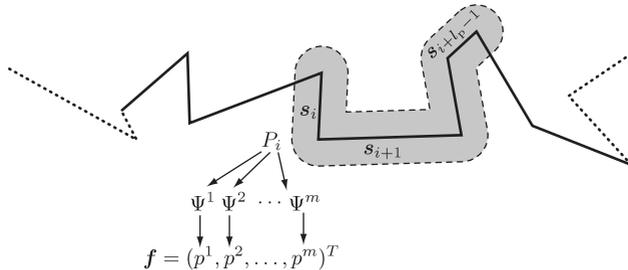


Figure 1.5: Extraction of features from a trajectory part.

The elements of the feature vector should be informative with respect to different properties of the sequence of steps. The operator Ψ^1 for computing the *net displacement* in a trajectory part P_i of l_P steps, for example, computes the overall directionality of motion as:

$$\Psi^1 : P_i \rightarrow p_i^1, \quad p_i^1 = \|\mathbf{x}_{i+l_P} - \mathbf{x}_i\|. \quad (1.9)$$

It is important that all features are invariant with respect to rigid-body rotations and translations of entire trajectory parts.

On the one hand, designing the features is often guided by prior knowledge about the physical process that generated the motion patterns. For example, motor-driven transport along intra-cellular filaments should be somewhat fast and rather straight. On the other hand, features can be purely descriptive or reflect a subjective impression of what is characteristic for a given motion pattern. For example, objects undergoing diffusive motion may seem to visit areas that are roughly elliptic in shape¹. Fi-

¹By, for example, inspecting the density of a brownian bridge [14] pinned at the end-points of the diffusion one can come up with more precise statements than this.

nally, one can use any scalar function of the positions \mathbf{x} as additional feature. This has the advantage that any information in the training set can be exploited that may help classify the data. One risks, however, adding uninformative functions to the feature set, which may complicate classification and increase the computational cost.

The collection of features is highly specific to each motion pattern. As an alternative to an expert’s choice (or educated guess), one can also perform systematic *feature selection* on a large set of features to determine an “optimal” feature set. Similar to the cross-validation strategy followed for tuning the SVM parameters (Section 1.3.2.3), one starts from a potentially very large set of features and selects the subset that yield best classification performance. Depending on the size of the original feature set and the cost of cross-validation, the search through all combinations of features may be performed exhaustively, or rely on some search heuristic.

Step 3: Feature classification. Feature classification evaluates whether a feature vector \mathbf{f}_i computed from a trajectory part P_i corresponds to a certain class c of motion or not, thus constituting a binary classification problem. I use C -SVMs [127] as classifiers. The SVM assigns each feature vector an output $o_c \in \{0, 1\}$. An output value of 1 indicates that the steps in the trajectory part are believed by the classifier to belong to class c , while the opposite is true for an output of 0.

Step 4: Classifier output filtering and class label assignment. The present feature extraction and classification approach raises the following problems: First, due to the overlap of the trajectory parts when computing features, each step is classified multiple times and may be assigned contradictory classification outputs with respect to a certain pattern. The final goal, however, is to give each step a unique label. Second, the outputs of the classifiers for the different patterns may suggest the affiliation of a step to more than one of the different patterns, necessitating a decision between classes. Third, additional conditions such as a minimum length of a trajectory segment may have to be satisfied. Although the features are computed in trajectory parts of a given length, this does not imply that all steps in that part are assigned the same class label. Without any regularization, trajectories could thus be *over-segmented* into insignificant fragments that are of little value for subsequent statistical analysis.

These ambiguities are resolved by post-processing the outputs of the different pattern-specific classifiers. For each pattern c , features are computed in overlapping trajectory parts and subsequently classified. This results in a series $O_c = \{o_{c,i}\}_{i=1}^{n-l_P}$ of classifier outputs for each pre-defined pattern c (if trajectories have been preprocessed, n has to be replaced with \tilde{n} to account for the potentially reduced number of steps). Each output is interpreted as a vote for the class-membership of all steps within the corresponding trajectory part. Since the trajectory parts overlap, each step gets multiple votes from the different pattern-specific classifiers. Based on these votes, each step k in the trajectory is assigned a score $s_{c,k}$ that represents the likelihood that this step belongs to pattern c . Ignoring the order of votes, a weighted sum of all outputs that are relevant for a particular step (Fig. 1.6B) is computed:

$$s_{c,k} = \sum_{j=1}^{l_P} a_{c,j} o_{c,k-j+1} \quad k \in [1, n-1] \quad o_{c,i} = 0 \quad \forall i < 1, \quad (1.10)$$

which amounts to a discrete convolution of the output sequence O_c with a filter kernel $A_c = \{a_{c,1}, \dots, a_{c,l_P}\}$:

$$\{s_{c,k}\} = O_c * A_c. \quad (1.11)$$

This representation is purely conceptual and is intended to highlight the connection of this step to related tasks in, for example, time-series analysis. The convolution kernel allows encoding additional prior knowledge or constraints. An ideal kernel would produce jumps in s whenever the true motion pattern changes, while keeping s constant otherwise. In that sense, the convolution operation is an attempt to do *inverse filtering* in order to reconstruct the true state (which motion pattern is present) from a noisy (due to misclassifications) and blurred (due to the trajectory parts extending over pattern transitions) signal. In general this is a hard problem. In the present work I use a uniform box kernel, corresponding to setting all weights $a_i = 1/l_P$, thus equally distributing the outputs among all steps within a trajectory part. Designing more sophisticated kernels might improve the overall performance of the trajectory segmentation algorithm.

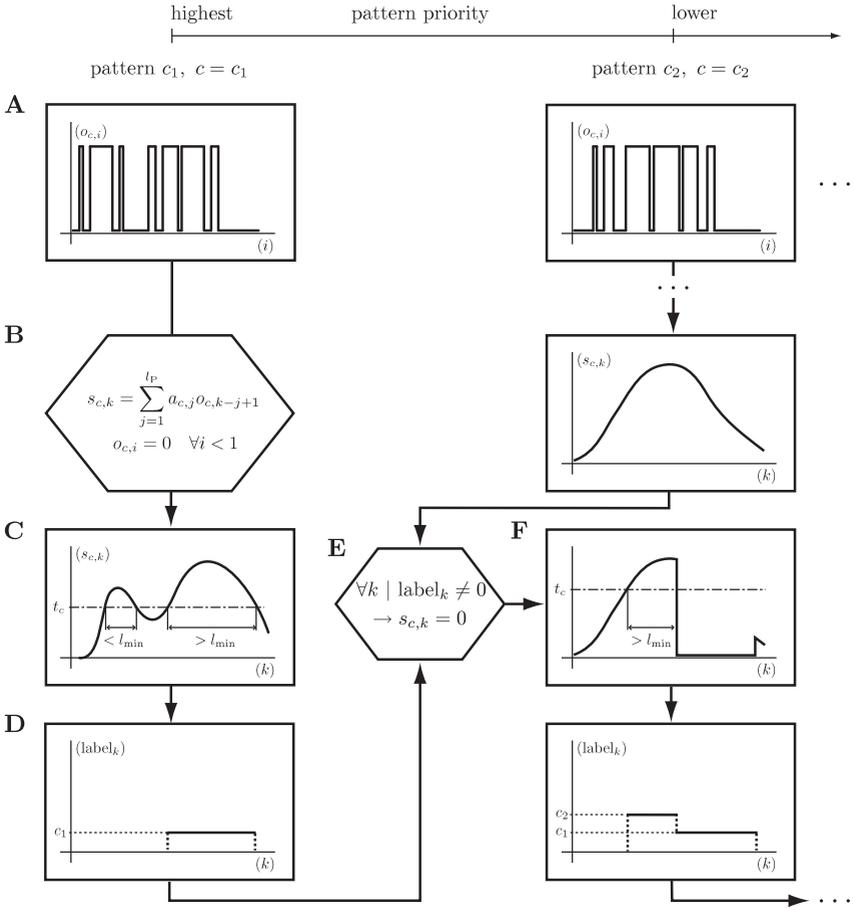


Figure 1.6: Algorithm for class label assignment using multiple sequences of binary classifier outputs $o_{c,j}$. The mapping of labels from steps of a filtered trajectory to the original unfiltered trajectory is not shown.

As stated before, the convolution in Equation 1.11 does not yield an ideal sequence of scores. Spurious classifications of trajectory parts might add a random component to the sequence of outputs O_c , which is reflected in a reduced difference between signal (presence of the pattern) and background (absence of the pattern) in the scores s . The problem of assigning the correct label is thus equivalent to finding regions in the score sequence s that are higher than the average score level caused by spurious classifications. A straightforward solution is to use a threshold filter, thus generating a binary signal (either 0 or 1) (Figure 1.6C).

The choice of the threshold t_c is constrained by the requirements of the application. A low value may be beneficial for precise localization of transitions between patterns. A lower bound is given by the expected rate of false positive classifications. High thresholds improve the separation of short segments of the same pattern and reduce false detections, while increasing the risk of missing some segments. An upper bound for the threshold is given by the expected rate of true positive classifications.

So far, it has not been discussed how to deal with the situation when a step has high scores for multiple patterns. As discussed in Section 1.3.2.2, the solution to this multi-class classification problem is to use priorities among the patterns. Starting from the pattern with highest priority, the classification output is transformed into a sequence of scores and the threshold filter is applied. In order to reject short and statistically insignificant segments it is required that the binary signal equals 1 for at least l_{\min} steps. For the steps in the trajectory that meet these constraints the binary signal is then converted to a sequence of step class labels by multiplication with c (Figure 1.6D). This process is repeated with all patterns in the order of descending priority. The scores of all steps that have already been associated with a pattern are set to zero in further iterations (Figure 1.6E and F), which effectively prohibits multiple class associations.

All steps that retain a final label of zero are considered not classified. If the trajectory was filtered prior to feature extraction, the label sequence has to be mapped back onto the steps of the original trajectory (not shown in Figure 1.6).

1.3.3 ADAPTATION FOR SEGMENTING VIRUS TRAJECTORY DATA

So far, no specific motion patterns, trajectory features, and options for the various filters have been specified. The trajectory segmentation algorithm will now be tailored to segmenting trajectories of single human Adenovirus serotype 2 (Ad2) particles moving inside or on the plasma membrane of a range of human cell types.

Ad2 enters cells via receptor-mediated endocytosis and is rapidly released into the cytoplasm (for a review see [102]). It requires an intact microtubule cytoskeleton to transport its DNA genome into the nucleus. Initial quantitative analyses of viral motion in the cytoplasm revealed long-range directional transport in normal cells, and short-range transport events in cells treated with the microtubule-depolymerizing drug nocodazole (reviewed in [51]).

Motion of Ad2 particles is observed by time-lapse microscopy at an image acquisition frequency of 50 Hz. A semi-total internal reflection fluorescence microscopy (semi-TIRFM) was used for acquisition of trajectories of intra-cellular virus motion. Standard TIRFM allowed limiting observations the virus particles on the plasma membrane. TIRFM achieves a specificity to the sample surface by restricting the illumination to a narrow region above the glass-sample interface [144]. Videos were processed with a single-particle tracking program [125], yielding the 2D trajectories of single Ad2. The signal-to-noise ratio of the image sequences was 2.5, yielding a standard deviation of the measured virus positions of 20 nm [125], referred to as *position noise*.

Visual inspection of Ad2 trajectories indicates that they contain different patterns of motion with a range of time scales (see segmented example trajectories in Figure 1.9). Motion frequently alternates between different patterns. Ad2 trajectories are, therefore, an ideal case to demonstrate the performance of the present trajectory segmentation algorithm. The focus is on the analysis of patterns that exhibit a strong non-random nature and that may thus be related to specific biophysical mechanisms. On very short time scales, position noise masks much of the structure of the motion, thus making it hard to distinguish random and non-random motion patterns there.

1.3.3.1 Motion Patterns in Ad2 Trajectories

Four patterns of Ad2 motion can be visually distinguished:

- Confined motion,
- slow drift,
- fast drift, and
- fast directed motion.

Confined motion is characteristic of receptor-ligand complexes experiencing steric hindrance [82]. Slow drifting motion of filopodial actin filaments is powered by myosin motors and actin dynamics as observed in cultured neuronal cells [101, 85]. Fast drifting and directed motion typically depends on microtubules and motor proteins [146, 51]. Between phases of either of the four patterns, Ad2 motion often appears random, which could correspond to (free) diffusion. Nevertheless, no attempt is made to identify diffusion as a separate motion pattern. Instead, trajectory parts that do not fall into any of the four classes are considered not classified. This allows performing four sequential binary classifications against this fifth auxiliary class, as discussed in Section 1.3.2.2.

1.3.3.2 Characteristic Features of Ad2 Trajectories

In the following, the trajectory features used for the classification of the motion patterns described in Section 1.3.3.1 are defined. These features constitute the set union of the pattern-specific feature sets as found by systematic feature selection on a larger set (see Step 2 in Section 1.3.2 and Section 1.3.3.5).

CHAPTER 1. QUANTIFYING THE DYNAMICS OF POINT-LIKE INTRA-CELLULAR OBJECTS

Some of the features are motivated by specific physical motion models. Others, however, are ad-hoc definitions. The names of the seven features that were selected in at least on pattern-specific feature set are:

1. Net displacement,
2. straightness,
3. bending,
4. efficiency,
5. asymmetry,
6. point position skewness, and
7. point position kurtosis.

The operator Ψ^1 for computing the *net displacement* for a trajectory part P_i of l_P steps is given by:

$$\Psi^1 : P_i \rightarrow p^1, \quad p^1 = \|\mathbf{x}_{i+l_P} - \mathbf{x}_i\|. \quad (1.12)$$

Straightness and *bending* are measures of the average direction change between subsequent steps.

Straightness is defined as

$$\Psi^2 : P_i \rightarrow p_i^2, \quad p_i^2 = \frac{1}{l_P - 1} \sum_{j=i}^{j+l_P-2} \cos \beta_j \quad (1.13)$$

and the operator for *bending* is

$$\Psi^3 : P_i \rightarrow p_i^3, \quad p_i^3 = \frac{1}{l_P - 1} \sum_{j=i}^{i+l_P-2} \sin \beta_j. \quad (1.14)$$

In both cases, β_j denotes the signed angle change between step \mathbf{s}_j and \mathbf{s}_{j+1} with changes to the left being attributed a positive sign.

Efficiency relates the squared net displacement to the sum of squared step lengths:

$$\Psi^4 : P_i \rightarrow p_i^4, \quad p_i^4 = \frac{\|\mathbf{x}_{i+l_P} - \mathbf{x}_i\|^2}{l_P \sum_{j=i}^{i+l_P-1} s_j^2}. \quad (1.15)$$

Asymmetry was previously proposed as a feature for the detection of directed motion [65]. The 2D equivalent, accounting for non-cylindrically symmetric point distributions is used, thus:

$$\Psi^5 : P_i \rightarrow p_i^5, \quad p_i^5 = -\log \left(1 - \frac{(\lambda_1 - \lambda_2)^2}{2(\lambda_1 + \lambda_2)^2} \right). \quad (1.16)$$

λ_1 and λ_2 are the eigenvalues of R , the 2D gyration tensor of the set of all points $\mathbf{x} = (x_i, y_i)^T \in P_i$:

$$R = \sum_i \begin{pmatrix} x_i^2 & -x_i y_i \\ -x_i y_i & y_i^2 \end{pmatrix}. \quad (1.17)$$

Position *skewness* and *kurtosis* measure the asymmetry and peakiness of the distribution of points in P_i , respectively. For their calculation, the positions $\mathbf{x}_j \in P_i$ are projected onto the dominant eigenvector \mathbf{v} of R , yielding scalars $x_j = \mathbf{x}_j \cdot \mathbf{v}$. *Skewness* and *kurtosis* are defined as the third and fourth moment of the set of x_j derived from the trajectory part P_i , respectively, thus:

$$\Psi^6 : P_i \rightarrow p_i^6, \quad p_i^6 = \frac{\sqrt{l_P + 1} \sum_{j=i}^{i+l_P} (x_j - \langle x_j \rangle)^3}{\left(\sum_{j=i}^{i+l_P} (x_j - \langle x_j \rangle)^2 \right)^{3/2}} \quad (1.18)$$

and

$$\Psi^7 : P_i \rightarrow p_i^7, \quad p_i^7 = \frac{(l_P + 1) \sum_{j=i}^{i+l_P} (x_j - \langle x_j \rangle)^4}{\left(\sum_{j=i}^{i+l_P} (x_j - \langle x_j \rangle)^2 \right)^2}. \quad (1.19)$$

Apart from the features defined above, additional features were considered, but none of them was ever selected by the systematic feature selection.

Position noise complicates robust detection of drifting motion since the net displacement is very small on small time scales. The noise can be partly removed by means of a trajectory filter function Φ .

Fast drifting motion only becomes apparent on time scales on the order of seconds, while slow drifting motion is roughly one order of magnitude slower. I therefore use the box filter (Equation 1.8) with widths $w = 10$ and $w = 50$ to generate smoothed and down-sampled trajectories \tilde{T} for the detection of fast drifting motion and slow drifting motion, respectively. This amounts to averaging 0.2 and 1.0s of the original trajectory per new position measurements. Detection of fast directed motion does not require filtering of the trajectories, since the typical displacement per step is comparable to or larger than the position noise.

On very short time scales, the confined motion might look like free diffusion, or any other motion type. The confinement character only becomes apparent when the expected net displacement of the unconfined motion exceeds the size of the region to which the virus particle is confined. Confined motion is therefore better characterized by a typical length-scale of confinement, rather than a time-scale as done with directed and drifting motion types. The apparent size of the regions of confinement is comparable to the magnitude of the position noise. No trajectory filter function is therefore used for the detection of confined motion.

Features are extracted from trajectory parts of lengths l_P , chosen separately for each motion pattern. The values reflect assumptions about the velocities and typical lengths of the different motion patterns. In a similar way, the minimum lengths l_{\min} of the segments of the different motion patterns are chosen. The thresholds t_c for transforming the classification scores s into class labels are manually selected to yield a good balance between sensitivity and specificity of pattern detection on a *per segment* basis (values t_c , l_{\min} , and l_P are given in Table 1.1). Therefore, the ability to correctly identify the correct sequence of segments was optimized, instead of optimizing for correct labeling of individual steps.

Finally, the priorities among the patterns need to be defined. Since three

of the four patterns correspond to more or less directed movements, the patterns can be ranked according to their typical velocities (see Table 1.1). That is, directed motion is detected first, followed by fast and slow drifting motion, and, finally, confined motion. This effectively avoids misclassification of, for example, long stretches of directed motion as fast drifting motion.

	priority	w	l_P	t_c	l_{\min}
Directed	1	–	25	0.15	20
Fast drift	2	10	10	0.5	8
Slow drift	3	50	10	0.3	8
Confined	4	–	50	0.3	30

Table 1.1: Parameters used in the trajectory filters and classification output filters. Note that for fast and slow drifting motion the units of the lengths l_P and l_{\min} are given number of steps of the respective filtered trajectories \tilde{T} .

1.3.3.4 Training Data

The most important ingredient for a successful adaptation of the trajectory segmentation algorithm is the training set. Using supervised classifiers allows one to *implicitly* define the input–output behavior of the segmentation algorithm through example data. The definition of informative features is then a prerequisite to exploit the structure in the example data. This is in contrast to the different pre- and post-processing filters, which are ad-hoc solutions to technical problems of the classification approach.

Training data could come from different sources. In the virus motion example an expert manually segmented example trajectories. Alternatives include synthetic trajectories based on either statistical models for the structure of the data or biophysical models of virus particle motion.

Here, training feature vectors are computed in windows randomly placed over a manually segmented trajectory. These feature vectors are given

class labels according to the most abundant motion pattern in the trajectory part covered by the window. The initial training set can later be augmented by samples from manually segmented trajectories where the *automatic* segmentation has previously failed. By cycling through tests and manual segmentation, the training set is iteratively improved such that it includes a large number of samples that are informative with respect to the decision boundary between classes. The sizes of the resulting training sets are given in Table 1.2.

Failure to achieve reasonable segmentation of trajectories after including them in the training set indicates a deficient feature set or ambiguous definitions of the motion patterns by the expert. For certain motion patterns it may thus be favorable to use synthetic training data. One example are motion patterns that do not last long and that are hence hardly visible in real trajectories. In such cases, synthetic data can be used to train the classifier. This does, however, require that a trustworthy model of the sought motion type is available. Another option is to combine synthetic and real trajectory data in the training set. In this case the synthetic data provide ideal (in the sense of the hypothetical motion model) positive samples, while the real data account for the true noise level and deviations from the model, as well as the potentially non-trivial distribution of negative samples in feature space.

1.3.3.5 Feature Selection and SVM Parameter Optimization

The minimal set of features that is suited to discriminate between the two classes of a given set of training samples is sought. A small number of features is desirable for two reasons: it decreases the computational cost of training and classification, and it yields a more robust classifier by reducing the complexity of the decision boundary. Given the training sets, SVMs are trained and tested using all possible combinations of a set of features (including, among others, all features introduced in Section 1.3.3.2). Sensitivity and specificity (see Section 1.3.2.4) are used to measure the generalization capability. They are estimated by five-fold cross-validation (see Section 1.3.2.4) on the training set. For each feature set size, the set with the best performance is selected, where performance is measured as

the average of sensitivity and specificity. As small sets are favored, which is not significantly outperformed by any larger one is chosen. This procedure is carried out independently for each pattern.

The selected feature sets for the four pre-defined motion patterns are:

- Fast directed motion: $\{p^1, p^2, p^3, p^4, p^5, p^7\}$;
- fast drifting motion: $\{p^1, p^2, p^3, p^5, p^6\}$;
- slow drifting motion: $\{p^1, p^2, p^4\}$;
- confined motion: $\{p^1, p^2, p^3, p^5, p^6, p^7\}$.

None of the tested additional features besides the seven described in Section 1.3.3.2 was ever chosen.

Given these sets of features, the SVM parameters $\{C, \sigma\}$ are adjusted by exhaustive search on a grid. As before, classification performance is measured in terms of sensitivity and specificity, estimated by five-fold cross validation on the training set. In principle, feature selection and optimization of the SVM parameters should be done simultaneously. Due to significantly higher computational cost, however, this approach is not practical and is thus not followed. Instead, the simpler sequential procedure described above is performed. The training set sizes for all motion patterns, the performance of the respective classifiers, and their optimal parameter values are summarized in Table 1.2.

1.3.4 EVALUATION

The trajectory segmentation algorithm is tested on trajectories where ground truth is known. Three tests are performed: The first determines to what extent the pre-defined patterns of Ad2 motion are found in trajectories of purely diffusive motion. Except for confined motion, which can be reproduced by extremely sub-diffusive behavior, none of the proposed motion patterns is compatible with a diffusion model (see Section 1.3.3.1). Therefore, the number of detected segments should not exceed the expected small random base level. The second test assesses the detection

	C	σ	N_{pos}	N_{neg}	Sens.	Spec.	Avg.
Directed	35.94	0.774	274	1370	0.996	0.994	0.995
Fast drift	12.92	3.594	338	1690	0.897	0.985	0.941
Slow drift	100	0.464	375	1082	0.859	0.939	0.899
Confined	4.46	0.464	2031	2039	0.975	0.956	0.966

Table 1.2: Summary of the SVM parameters used, the achieved classification performance on Ad2 data, and the parameters used in the classification output filters. N_{pos} and N_{neg} denote the number of positive and negative training samples, respectively; Sens.: sensitivity, Spec.: specificity, Avg.: (Sens. + Spec.)/2.

power on synthetic heterogenous trajectories that are designed to resemble real Ad2 trajectories. The third test illustrates how segment length relates to the detectability of a motion pattern, exemplified on stretches of fast directed motion that is preceded and followed by diffusive motion.

1.3.4.1 Generation of Synthetic Trajectories

Motion of virus particles can be described by a stochastic process that generates random positions $\mathbf{x}(t)$ according to a transition density $p(\mathbf{x}, t + \delta t | \mathbf{x}_0, t)$. For free (unrestricted) isotropic diffusion of a particle in d dimensions the transition density is given by [14, 1]

$$p(\mathbf{x}, t + \delta t | \mathbf{x}_0, t) = \frac{1}{(4\pi D\delta t)^{d/2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_0\|^2}{4D\delta t}\right), \quad (1.20)$$

which is a d -dimensional isotropic Gaussian with mean $(0, \dots, 0)^T$ and identical variance $2D\delta t$ in all dimensions. This means that the components of the displacement $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0$ are uncorrelated. Consequently, the right-hand-side of Eq. 1.20 can be factored into d 1-dimensional Gaussians:

$$p(\mathbf{x}, t + \delta t | \mathbf{x}_0, t) = \prod_{i=1}^d \frac{1}{(4\pi D\delta t)^{1/2}} \exp\left(-\frac{\delta x_i^2}{4D\delta t}\right), \quad (1.21)$$

where δx_i is the i -th component of $\delta \mathbf{x}$. Eq. 1.21 forms the basis of a simple algorithm for simulating diffusion of virus particles: For a given time increment δt the particle is displaced from \mathbf{x}_0 by $\delta \mathbf{x}$, where the elements of the displacement vector are independent random numbers drawn from a normal distribution with mean 0 and variance $2D\delta t$.

In order to generate trajectories with sub-diffusive motion, the Weierstrass-Mandelbrot function [11, 123, 151]

$$W_i(t) = A \sum_{n=n_{\min}}^{n_{\max}} \frac{1}{\gamma^{nH}} (\cos U_i(n) - \cos [\gamma^n t^* + U_i(n)]) \quad (1.22)$$

is sampled independently for all dimensions i at discrete times t . $t^* = 2\pi t/t_{\max}$, with t_{\max} the total time of the synthetic trajectory. Following Saxton [123], I set $\gamma = \sqrt{\pi}$, $n_{\min} = -8$, and $n_{\max} = 48$. The $U_i(n)$ are uniformly distributed random numbers between 0 and 2π . Computing the cumulative sum of $W_i(t)$ for each dimension yields the point sequence $\mathbf{x}(t)$ of the trajectory. Trajectories of any diffusion type (as characterized by the diffusion constant D and the MSS slope (see Section 1.2) can be generated by adjusting the parameters H and A to match the desired behavior. Confined motion is simulated by adjusting H such that the MSS slope is 0.1, resulting in strong visual similarity to the noisy trajectory of an immobile particle. Directed motion, fast drift, and slow drift are created by superimposing bent directed motions of different speeds (300-600, 20-200, and 8-16 nm/s) and durations (0.5-4, 2-15, and 10-50 s) onto trajectories of confined motion.

Heterogeneous synthetic trajectories are created by concatenating segments of homogeneous trajectories of different motion types.

1.3.4.2 Segmentation of Synthetic Trajectories

In the first test, homogeneous synthetic trajectories are used to estimate the probability of the four motion types to occur in uniform random walks. This provides confidence intervals for assessing the statistical significance of the Ad2 results. Two types of homogenous trajectories are used: unrestricted isotropic diffusion and isotropic sub-diffusive motion. Global

trajectory statistics of the synthetic trajectories, such as length, diffusion constant D , and MSS slope (only for synthetic sub-diffusion) were matched with the values of a collection of Ad2 trajectories on the plasma membrane of M21 cells.

The segmentation results are summarized in Table 1.4. In synthetic trajectories containing only normal isotropic diffusion, 98.9% of the steps have correctly been labeled as not classified, and only minor amounts of the four pre-defined motion patterns were found. For sub-diffusive synthetic trajectories, the segmentation algorithm yielded a comparable fraction of confined motion (67.2%) as in the Ad2 trajectories (55.9 to 65.5%, Table 1.4), while all other motion types were extremely rare (less than 0.1%). This confirms that detection of any of the directional motions (fast and slow drift, directed) is highly significant with respect to a null model of purely isotropic diffusive motion. Furthermore, confined motion is not a random artifact in unrestricted isotropic diffusion, but compatible with a sub-diffusion model with very small MSS slope.

The second test focuses on heterogenous synthetic trajectories. Two representative results from the segmentation of heterogeneous synthetic trajectories are shown in Figures 1.7A and 1.7B. The automated segmentation is visually convincing, and practically all sufficiently long segments were, at least partially, detected.

In order to quantify the quality of the segmentation, the detection sensitivity and the specificity is measured on a per-step basis for each motion pattern. This is done on a large set ($N = 200$) of long (~ 4000 steps) synthetic heterogeneous trajectories. The results are given in Table 1.3. In all cases, the rates of correct classification (sensitivity) were $>90\%$ with rates of correct rejection (specificity) $>95\%$. The largest fraction of spurious positives (13%) is observed for fast drifting motion, which has, however, a low frequency (1.5%). The largest contribution to the errors comes from transition points between different motion patterns. While such transitions were always correctly *detected*, they were not always precisely *localized*. Within longer stretches of one motion pattern, almost no classification errors were observed. This confirms that segments as a whole are trustworthy, that is, they are hardly corrupted by over- or under-segmentation artifacts. This property is crucial for the quantification of extensive prop-

1.3. TRAJECTORY SEGMENTATION

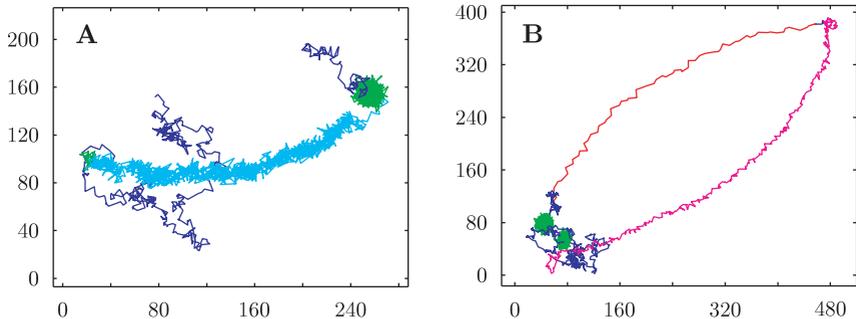


Figure 1.7: Segmented synthetic trajectories mimicking Ad2 motion in M21 cells. Segments are color coded. Green: confined motion; cyan: slow drifting motion; magenta: fast drifting motion; red: directed motion; blue: not classified.

	Number of steps	Sensitivity	Specificity
Directed	2921	91.3 %	99.9 %
Fast drift	8805	97.4 %	99.8 %
Slow drift	133929	94.5 %	99.4 %
Confined	311218	95.4 %	97.2 %
Not classified	142944	95.1 %	96.4 %

Table 1.3: Detection rates for the different motion patterns in synthetic heterogeneous trajectories computed on a per-step basis.

erties of segments, such as their length.

Detection is harder for shorter segments. In order to assess the capability of the method to detect short segments, synthetic heterogeneous trajectories of random walks with intermediate segments of directed motion of varying length are analyzed. A segment of directed motion is considered detected if at least 80% of its steps are assigned the correct class label. The result is shown in Figure 1.8. Above a length of about 25 steps, the detection rate quickly approaches 100%. Note that 25 in the present case steps correspond to 0.5s of virus motion, which is comparable to

typical run-lengths [152, 44] of motor-driven transport along filaments of the cytoskeleton. Therefore, the trajectory segmentation algorithm should be capable of providing statistics of motor-driven transport, such as run-length and velocity distributions [74, 108, 133], from in vivo experiments using a great variety of cargos.

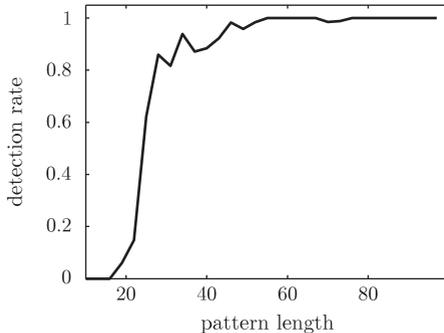


Figure 1.8: Detection rate of directed motion as a function of the true segment length. Only segments with $>80\%$ correct class labels are counted as detected. $N = 200$ trajectories.

1.3.5 APPLICATION

The trajectory segmentation algorithm is applied to real trajectories of Ad2 on live cells. The example results shown in Figure 1.9 illustrate that the algorithm properly discriminates nearby segments of confined motion (Figure 1.9B) and correctly identifies slow drifting motion (Figure 1.9C), fast drifting motion (Figure 1.9B), and directed motion (Figure 1.9A and D).

A large number of intra-cellular trajectories of Ad2 from semi-TIRF recording conditions was analyzed, either in the absence ($N = 4388$) or presence ($N = 2399$) of the microtubule inhibitor nocodazole. The frequencies of both fast drift and directed motion are strongly reduced in the presence of nocodazole, namely from 2.07% to 0.17%, and from 0.58% to 0.06%,

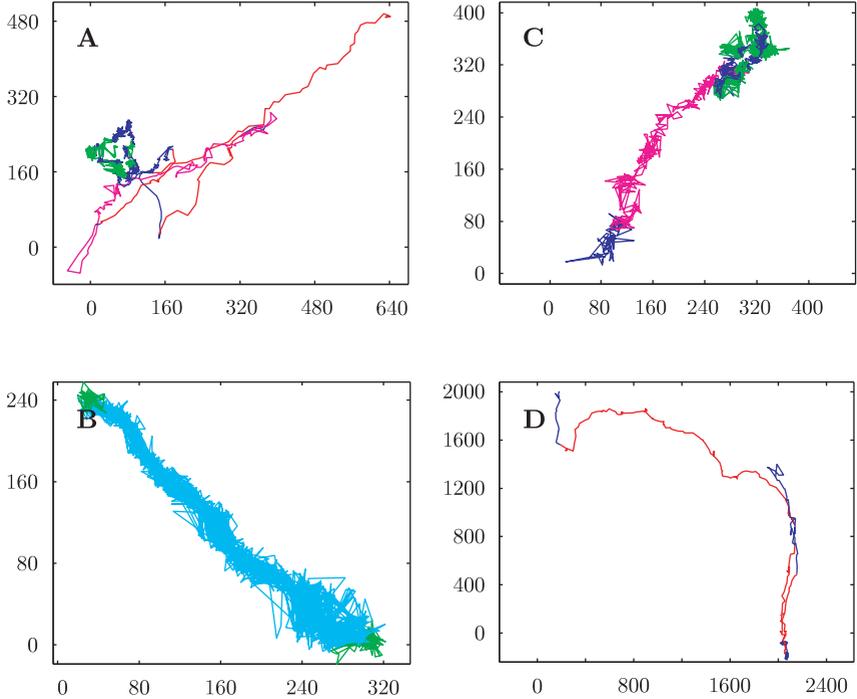


Figure 1.9: Segmented intra-cellular trajectories of Ad2 in M21 cells. Segments are color coded. Green: confined motion; cyan: slow drifting motion; magenta: fast drifting motion; red: directed motion; blue: not classified. (A and B) Examples of strongly heterogeneous Ad2 trajectories composed of directed motion (not present in B), fast drifting motion, confined motion, and non-classified motion. (C) A segment of slow drifting motion extends over almost the entire trajectory. (D) Example of a trajectory with several separate phases of directed motion. Distances are in units of nanometers.

respectively (Table 1.4). Besides drastically reducing the frequency of fast drift and directed motion, nocodazole increased the fraction of confined motion from 54.1% to 62.7%, while the amount of slow drift was reduced by 50%. Fast drift and directed motion of intra-cellular Ad2 thus depend on intact microtubules, possibly involving motor proteins such as dynein-dynactin or kinesins [142, 146], or microtubule dynamics [31, 36].

In the standard model of Adenovirus infection, most virus particles cluster around the nucleus one hour after infection. This accumulation crucially depends on intact active transport. Surprisingly, however, intra-cellular trajectories in our data show only small fractions of microtubule-dependent transport types (directed motion and fast drift). Does this contradict the standard model? While active transport is indeed surprisingly rare, it is responsible for most of the net transport distance toward the nucleus. Some 30 s of directed motion suffice to travel cell-scale distances ($25\ \mu\text{m}$). Diffusive motion and confinement are much more frequent than active transport. Nevertheless, they are less relevant for net virus transport to the nucleus: Empirical diffusion constants for virus particles are on the order of $10^{-3}\ \mu\text{m}^2/\text{s}$, which yields net displacements on the order of only a few μm per hour.

Although the small observed fraction of directed motion may be sufficient for effective virus infection, it could still be underrepresented in the segmented trajectories for at least three reasons: First, the cells are not ideally flat. Since semi-TIRF microscopy does not visualize viruses far away from the glass-water interface, vertical motion may be missed altogether. Second, trajectories may be incomplete or corrupted by spurious linking, which is more likely to happen for fast moving particles and high densities of moving particles. Third, directed motion can be of very short duration and hence hard to detect. Stochastic unbinding of motor proteins from filaments typically yields an exponential distribution for the duration of active transport. The unbinding rate is on the order of seconds, which means that a non-negligible amount of directed motion is not detected due to its insufficient duration (see Figure 1.8). Furthermore, the motion statistics are computed relative to the set of all detected virus particles. This set includes a significant fraction of viruses that have failed to enter the cell or are immobilized on the glass. Relative changes in the fractions reported in Table 1.4 may thus be more informative than absolute values.

A further application of the trajectory segmentation algorithm considers virus motion on the surface of different cell types within the first 15 minutes after viral inoculation, including M21 human melanoma cells ($N = 414$), human embryonic kidney 293 cells ($N = 952$), human embryonic retinoblast 911 cells ($N = 403$), and primary human umbilical vein endothelial cells (HUVEC) ($N = 361$). Ad2 particles readily gained access to the basal side of the cells and were amenable to examination by TIRF microscopy.

As expected, there is a low fraction of at most 0.12% of directed motion in each cell type (Table 1.4). The fraction of slow drift varies between 6.6% and 14.9%, while the fraction of confined motion is between 57.1% and 65.5%. Not classified motion accounts for 21.4% to 32.3% of the trajectory steps. Remarkably, the fraction of slow drift is 3- to 7-fold larger than in trajectories of intra-cellular virus particles. The fractions of motion types and their standard deviations were estimated by bootstrapping [58]. Based on these estimates, the significance of the observed differences in the experimental conditions can be assessed. In the intra-cellular M21 data the standard deviations of the fractions of fast drift and directed motion are 0.21% and 0.06%, respectively, indicating that the nocodazole-mediated reduction of the observed patterns is statistically highly significant.

Significant differences are also observed for cell surface trajectories, most prominently in the fraction of slow drift and confined motion. The differences in directed motion are negligible due to the small total amount of this motion type. Together, these analyses suggest that the pattern frequencies of the different cell types constitute specific fingerprints. The differences in these fingerprints are significant, since the pattern fractions of the homogeneous synthetic data differ from the experimental data by more than 10 standard deviations. In conclusion the described motion types in both cell surface and intra-cellular Ad2 trajectories can not be explained by assuming that a simple random process underlies all trajectories. Instead, the data suggest a connection between trajectory segments and specific Ad2 interactions with cellular structures. The underlying cellular machinery possibly depends on the organization of the plasma membrane and the cytoskeleton.

	N	Directed		#	Fast drift		#	Slow drift		#	Confined		#	Not classified		#
		t [%]	t [%]		t [%]	t [%]		t [%]	t [%]		t [%]	t [%]				
<i>(A) Synthetic</i>																
Normal diffusion	3130	0.006 ± 0.003		6	0.36 ± 0.04		102	0.44 ± 0.09		25	0.33 ± 0.02		223	98.9 ± 0.1		3486
Sub-diffusion	3130	0.02 ± 0.01		16	0.04 ± 0.01		11	0		0	67.2 ± 1.0		6155	32.8 ± 1.0		9312
<i>(B) Intra-cellular</i>																
M21: No drug	4388	0.58 ± 0.06		209	2.07 ± 0.21		186	2.03 ± 0.38		39	55.9 ± 0.6		5549	39.4 ± 0.6		10305
M21: 50µM Noc.	2399	0.06 ± 0.02		19	0.17 ± 0.06		12	1.07 ± 0.40		10	65.4 ± 0.9		3048	33.3 ± 0.9		5476
<i>(C) Cell Surface</i>																
M21	414	0.10 ± 0.04		8	1.29 ± 0.31		26	9.69 ± 1.84		30	57.2 ± 2.0		821	31.7 ± 2.0		1270
293	952	0.003 ± 0.003		1	1.56 ± 0.24		92	14.9 ± 1.16		148	62.1 ± 1.4		2043	21.4 ± 0.8		3113
911	403	0.12 ± 0.06		8	2.05 ± 1.04		22	8.45 ± 1.87		21	57.1 ± 2.4		617	32.3 ± 2.0		1044
HUVEC	361	0.02 ± 0.02		1	1.05 ± 0.40		13	6.62 ± 2.04		14	65.5 ± 2.2		595	26.9 ± 2.0		965

Table 1.4: Frequency analysis of motion patterns in real Ad2 trajectories compared to synthetic data. (A) Synthetic homogeneous trajectories. (B) Intracellular motion patterns of Ad2 trajectories in unperturbed human M21 melanoma cells, and M21 cells treated with nocodazole (Noc.). (C) Cell surface motion patterns of Ad2 on human M21 melanoma, 293 epithelial, 911 epithelial, and primary HUVEC endothelial cells. Bootstrap estimates of standard deviations are indicated by \pm .

Segmentation of trajectories allows analyzing motion types independently, which is of particular importance for motion types that do not occur frequently and identification of transport models. Since they are tightly linked to the biophysical transport mechanism, the run-length and velocity of directed motion are of particular interest. For confined motion, the spatial extent of the area visited by the virus particle may characterize the mechanism that limits the mobility. Run-length and velocity distributions for the three directed motion types are summarized in Figures 1.10, 1.11, and 1.12. The duration and spatial extent of phases of confined motion are shown in Figure 1.13.

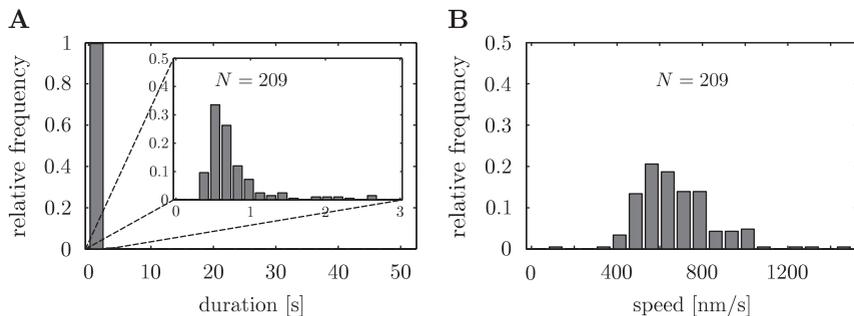


Figure 1.10: Histograms of parameters characterizing directed motion of Ad2 particles inside in M21 cells.

Fast drift typically lasts for 2 to 10 seconds with single-displacement speeds of 20 to 400 nm/s (Figure 1.11). Directed motion was faster (400 nm/s to >1000 nm/s), but lasted for only short periods of 0.3 s to 2.7 s (Figure 1.10). The end-to-end displacement of directed motion was nonetheless higher than the one of fast drifting motion. 49% of the fast drifting motion and 30% of the directed motion were present in tracks with overall MSS slopes below 0.5, suggesting a globally sub-diffusive process (see Section 1.2). This highlights the importance of trajectory segmentation for functional analysis. In addition, fast drift and directed motion were often saltatory and bidirectional (see Figure 1.9A and D), indicative of microtubule-based transport. In fact, the velocity distribution for directed motion peaks at 500 nm/s, which is of the same order as (but still less than) the na-

tive velocity of molecular motors such kinesin [143, 21, 9] or cytoplasmic dynein [146, 94, 112]. Surprisingly, some virus particles appear to move with higher average velocities than the native motor velocity. This may be a statistical effect, since the underlying trajectory segments can reflect less than 100 motor steps. An alternative explanation is that multiple motors work in parallel on the same cargo. In Chapter 4, a hypothetical mechanism is discussed that could explain how coupled motors mutually increase their velocities.

The shapes of the velocity and run-length distributions are very similar to those reported in references [74, 108, 133]. Such distribution have been explained with transport models of intra-cellular cargo, which allowed estimating important quantities such as the number of motors bound or forces exerted on the cargo, but also shed light on fundamental principles of cargo transport organization [108, 133, 44].

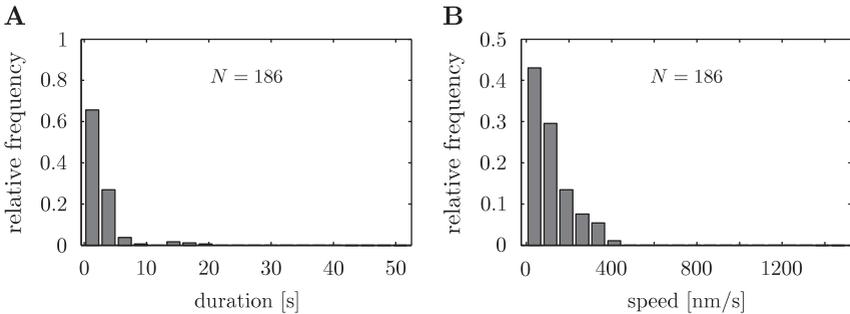


Figure 1.11: Histograms of parameters characterizing fast drifting motion of Ad2 particles inside in M21 cells.

Confined motion was typically short-lived with durations on the order of a few seconds within areas of on average 80 nm diameter (max. 150 nm) (Figure 1.13). This high degree of confinement is consistent with earlier observations [142]. It is possibly attributed to the viscous cytoplasmic environment and the dense cytoskeleton meshwork restricting cytoplasmic diffusion. Alternatively, it could be due to tethering of particles to immobile cellular scaffolds.

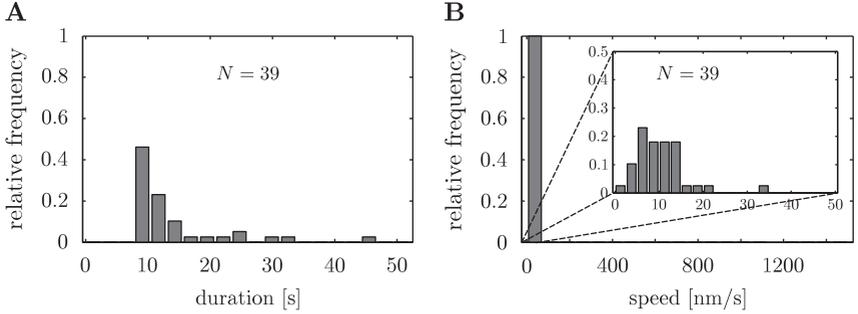


Figure 1.12: Histograms of parameters characterizing slow drifting motion of Ad2 particles inside in M21 cells.

Unlike confined motion, slow drifting motion proceeded without interruptions with speeds of 2 to 10 nm/s (Figure 1.12). These speeds were determined in time intervals of several seconds, comprising hundreds of displacement steps (Figure 1.12). A single step displacement was far smaller than the position noise of 20 nm. Slow drifting motion often lasted throughout the entire observation period, indicating a high persistence (see Figure 1.9C).

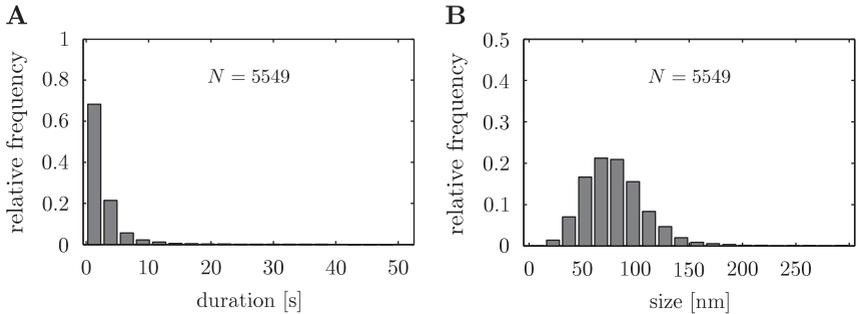


Figure 1.13: Histograms of parameters characterizing confined motion of Ad2 particles inside in M21 cells.

QUANTIFYING THE SHAPE AND DYNAMICS OF EXTENDED INTRA-CELLULAR OBJECTS

In this chapter I present a novel computational image analysis technique for precise and accurate segmentation of small intra-cellular objects from fluorescence microscopy images.

A central paradigm in structural cell biology is that the microscopic shape of sub-cellular objects is closely linked to their function. The shapes of involved objects can determine or constrain a biological function. Diffusion on membranes of organelles, for example, may be influenced by their complex morphology [126]. In eukaryotic cells membrane-bound organelles often serve as reaction compartments in which the cell maintains unique protein and lipid compositions [99]. On the length scale of organelles, the volume of the reaction compartment can be very important, as seen by its influence on steady state concentrations of even simple reactions. Organelles shapes are highly dynamics, due to, for example, fluctuations of internal constraints in membranes (e.g., recruitment of curvature-inducing lipids or proteins) or external forces (e.g., interactions with molecules of the cytoskeleton [98], forces from molecular motors [84, 133], fusion of

organelles [121, 120]). Organelle shapes can therefore be used as indirect readouts of many biological processes that are not yet well understood [109, 133]. Furthermore, correlating the positions and shapes of populations of different objects allows mapping cellular organization [110] and investigating specific interactions. Extracting shape information from live cells is therefore of great interest.

Advances in light microscopy [138] and the development of fluorescent tags for specific labeling of proteins [48] have rendered fluorescence light microscopy the most widely used experimental tool for imaging sub-cellular structures *in vivo*. The acquired images contain a wealth of information, which is, however, complex and under-explored, and requires reduction to a comprehensible form.

Visual interpretation of microscopy images by an expert is, and will remain, the gold standard for many applications. The reason is that the human brain is at extracting semantic information. Image analysis can be automated to some extent, but it is mostly limited to applications where one has *a priori* knowledge about the image content. The analysis is focussed on specific features of the images, which are assumed to correspond to physical properties of the studied objects in some way. The semantic information is given by the prior, for instance in form of a *structural model*, while the computer provides the quantitative information, that is, *parameter estimates* of the model. As opposed to humans, computers are objective. The possibility to process very large datasets increases confidence levels [92]. Furthermore, non-intuitive measurements can be performed with very high accuracy and precision. Microscopy data in cell biological studies have a number of properties that favor computers: The images are typically noisy, blurred, large, and the datasets can be large. Different biological conditions can yield small signals that would be missed in human inspection of a fraction of the available data.

In the past, the development of computational image processing tools to extract biologically relevant information from microscopy images has attracted great attention. A fundamental problem with microscopy images in cell biology is that the mapping from the observed objects to their image is not bijective, that is, information is lost in the imaging process. Depending on the size, shape, and spatial arrangement of the studied objects, this

loss of information can be severe, in which case the reverse mapping from images to objects is no longer intuitive. In the following sections I devise an image analysis algorithm that extracts information on the shape of small intra-cellular objects from fluorescence microscopy images. As demonstrated on synthetic and real image data, results are less biased than solutions based on classical image segmentation techniques. Before presenting the actual algorithm and example applications in Section 2.2, I review the fundamentals of fluorescence microscopy in Section 2.1.

2.1 MICROSCOPY, DECONVOLUTION, POINT SPREAD FUNCTIONS, ETC.

The outline reconstruction algorithm presented in this chapter is specifically tailored to images obtained by (confocal) fluorescence microscopy. Wide-field and confocal fluorescence microscopy are well-described by the laws of linear optics. In the following sections I review fundamentals of image formation in fluorescence microscopy and essential properties of fluorescent dyes, microscope optics, and detectors.

2.1.1 FLUORESCENCE IN CELL BIOLOGY

Cells produce very little contrast in standard bright-field microscopy because light in the visible spectrum interacts only weakly with cells or tissues. Absorption, which is usually the main source of contrast, is negligible for biological matter. That is, most of the light just passes unchanged from the light source through imaged cells onto the detector. The *phase* of light, however, is changed by the presence of biological matter. This effect is exploited in *phase contrast microscopy*, which allows observing living cells without having to introduce potentially harmful dyes.

In fluorescence microscopy contrast is generated by a change in *frequency*, or *wavelength*, of the light. The change in wavelength is generated by fluorescent molecules that absorb and emit light and different wavelengths (see Figure 2.1). The energy of a single photon is given by hc/λ , where h is Planck's constant, c the speed of light in vacuum, and λ the wavelength

of the photon. Absorption of a photon of the excitation wavelength brings the fluorophore to a higher energy level – the *excited state*. Some of the energy is dissipated without radiation. Then, the fluorophore jumps back to the *ground state*. In this process a photon is emitted that has a smaller energy, and hence larger wavelength, than the photon previously absorbed (Figure 2.1). The difference in the maxima of the absorption and emission spectra is called *Stokes shift*.

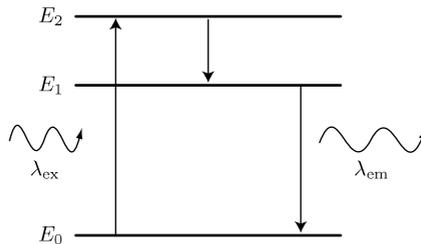


Figure 2.1: Jumps between energy levels in a fluorophore.

In fluorescence microscopy this effect is exploited in two ways: First, it is possible to shield the detector from excitation light that just passed through the sample. For this, one uses filters that selectively pass only a small range of wavelengths. Second, if one is able to selectively bind fluorophores to biological structures of interest, very specific observations can be made. The latter is typically achieved by fluorescent immunolabeling or by fluorescent *fusion proteins*. The trick is to introduce the gene sequence of a fluorescent protein in the sequence of the target protein, without hampering their respective functions. Translation of the combined sequence then creates chimeras of a target protein and a fluorescent protein that serves as a highly specific tag. The *green fluorescent protein* (GFP) and variants thereof are extremely popular fluorescent proteins [48].

2.1.2 BASICS OF IMAGE FORMATION

Image formation in fluorescence microscopy can be understood within the theory of *Fourier optics*. The central statement of the Huygens–Fresnel

principle is that each point of an unobstructed wave front is a source of a spherical wave. Light is an electromagnetic wave. The phase and frequency of the spherical electromagnetic waves are given by that of the original wave front under consideration. Consequently, the amplitude of the electromagnetic wave at any point beyond the source is a *superposition* of the spherical waves. This considers the waves' amplitudes *and* phases, that is, the waves produce an *interference pattern*, not just a superposition of *intensities*. From geometrical considerations (see Figure 2.2 for

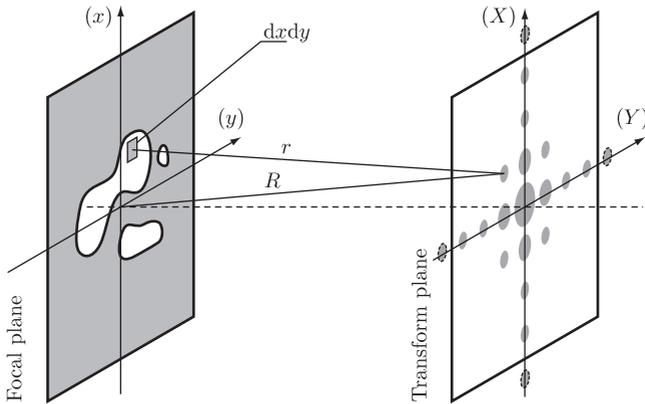


Figure 2.2: The spatial frequency decomposition of the aperture function (left plane) can be observed in the transform plane (right). Lenses between the planes are not shown for clarity.

coordinates) it follows that the electric field at the location (X, Y) in the *transform plane* is given by

$$E_{\text{tr}}(X, Y) = \frac{E_0 \exp(i(\omega t - kR))}{R} \iint_{\text{Aperture}} \exp\left(\frac{ik(Xx + Yy)}{R}\right) dx dy, \quad (2.1)$$

where E_0 is the amplitude, $k = 2\pi/\lambda$ the wave number, and $\omega = 2\pi c/\lambda$ the angular frequency of the spherical waves originating from the aperture. The *aperture function*

$$A(x, y) = A_0(x, y) \exp(i\phi(x, y)) \quad (2.2)$$

describes the spatial variation of the amplitude $A_0(x, y)$ and the phase $\exp(i\phi(x, y))$ of the point sources in the *focal plane* of the microscope objective (that is, the intensity and phase of the light emitted by the fluorophores of the sample that are in focus). Using the aperture function and ignoring the R -dependence of the amplitude, Equation 2.1 can be reformulated as:

$$E_{\text{tr}}(X, Y) \propto \iint_{-\infty}^{+\infty} A(x, y) \exp\left(\frac{ik(Xx + Yy)}{R}\right) dx dy, \quad (2.3)$$

and finally as:

$$E_{\text{tr}}(\kappa_x, \kappa_y) \propto \iint_{-\infty}^{+\infty} A(x, y) \exp(i(\kappa_x x + \kappa_y y)) dx dy, \quad (2.4)$$

where the *spatial frequencies*

$$\kappa_x = kX/R = k \sin \beta \quad (2.5)$$

$$\kappa_y = kY/R = k \sin \alpha \quad (2.6)$$

have been introduced. An important observation can now be made: Equation 2.4 describes the *Fourier transform* of the aperture function. In short notation:

$$E_{\text{tr}}(\kappa_x, \kappa_y) \propto \mathcal{F}\{A(x, y)\}. \quad (2.7)$$

One can thus observe the *power spectral density* of the aperture function in the transform plane (see Figure 2.2). This means that each point in the transform plane corresponds to a spatial frequency of the aperture function. The point's distance to the optical axis is proportional to the mapped spatial frequency. The intensity of light in each point is proportional to the weight of the corresponding spatial frequency in the frequency decomposition of the aperture function.

Up to this point no information is lost. If one was able to measure the interference pattern in the transform plane and perform an inverse Fourier transform, one would perfectly reconstruct the aperture function, that is, the imaged sample would be perfectly resolved.

In a microscope the light emitted from the sample is collected by the objective and focussed onto some detector through a series of optical ele-

ments. Magnification is of secondary interest, since its sole purpose is to ensure sufficiently fine sampling of the image on the detector. This detector can be a technical device such as a camera, consisting of an array of photo-sensitive elements, or the human eye. In the image plane, the waves corresponding to the different spatial frequencies of the aperture function again interfere. The pattern in the transform plane can again be imagined as sources of spherical waves. Therefore, the amplitude and phase of the electromagnetic wave in the image plane is the inverse Fourier transform of the electromagnetic field in the transform plane. Nevertheless, information is lost.

The reason for the information loss is that the optical device is not infinitely large. Only a small fraction of the light forming the pattern in the transform plane can actually be captured by the objective lens. With the diameter of the objective and the focal distance being fixed, a large amount of light will miss the objective at the side. This light corresponds to the high spatial frequencies of the aperture function. The pattern in the transform plane is hence cropped – it lacks the high frequency information (dashed circles in Figure 2.2). Therefore, the electromagnetic wave on the detector lacks the high spatial frequencies of the aperture function that were not captured. Mathematically, this lack of frequencies can be described by a weighting function $M_k(\cdot) \in [0, 1]$ in frequency space, where the subscript indicates the dependence on the wave number k . The electric field on the detector is

$$E_d(x, y) \propto \mathcal{F}^{-1}\{\mathcal{F}\{A(x, y)\} \cdot M_k(\kappa_x, \kappa_y)\}. \quad (2.8)$$

The *convolution theorem* states that the convolution of the two functions f and g is (under suitable conditions) the point-wise multiplication of the Fourier transforms of the two functions:

$$f * g = \mathcal{F}^{-1}\{\mathcal{F}\{f\} \cdot \mathcal{F}\{g\}\}. \quad (2.9)$$

Therefore, Equation 2.8 can be rewritten as

$$E_d(x, y) \propto A(x, y) * \mathcal{F}^{-1}\{M_k(\kappa_x, \kappa_y)\}. \quad (2.10)$$

The situation where $A(\cdot)$ is non-zero only in an infinitely small circular region is of particular interest. It corresponds to imaging a single *point*

source, which is described by the Dirac delta function $A(x, y) = \delta(x) \delta(y)$ (see Figure 2.3). The convolution of any function $f(x, y)$ with a Dirac delta centered at the origin is $f(x, y)$. The corresponding *irradiance* $I = \langle (\text{Re } E)^2 \rangle_t$ at the detector, that is, the image of a point source, is called *point spread function* P :

$$P \propto \left\langle \left(\text{Re} \left[\mathcal{F}^{-1} \{ M_k(\kappa_x, \kappa_y) \} \right] \right)^2 \right\rangle_t, \quad (2.11)$$

where $\langle \cdot \rangle_t$ denotes a time average. Using similar arguments as above, an analytic expression for the point spread function of an idealized microscope is derived in Reference [61].

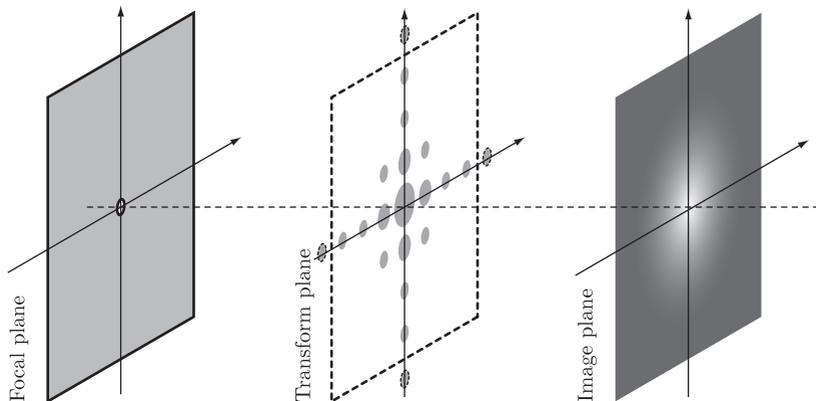


Figure 2.3: Due to the lack of high-frequency information (dashed circles in the middle plane), the image of the small point-like source (left plane) is blurred (right plane). Lenses between the planes are not shown for clarity.

Any $A(x, y)$ can be decomposed into a sum of Dirac point sources. For *incoherent* point sources, i.e., sources with uncorrelated phases, the *image* – and not just the electromagnetic wave – is a superposition of the images of the point sources, each of which being a point spread function. The aperture function can then be identified with the square root of the density of independent fluorophore molecules $O(x, y)$. Mathematically, the superpo-

sition of the single fluorophore images is described by the convolution:

$$I_d \propto O * P. \quad (2.12)$$

Strictly speaking, the above equation is only valid for a *linear* and *space-invariant system*, which is approximately true for a conventional fluorescence microscope. Fluorophores are perfect *self-luminous* point sources with uncorrelated photon emission. Therefore, they have uncorrelated phases and Equation 2.12 may be safely used to model image formation.

2.1.3 RESOLUTION AND BLURR

A key property of microscopes is their *optical resolution*. As outlined in Section 2.1.2, image formation in fluorescence microscopy is modeled as a convolution of the fluorophore density $O(\cdot)$ with the point spread function $P(\cdot)$. This smoothes out variations in $O(\cdot)$, i.e. the image appears blurred. The *Rayleigh* resolution limit refers to the minimum distance Δl between two point sources that still yields a bimodal image. Based on a model point spread function for an ideal fluorescence microscope, this limit can be expressed as

$$\Delta l = \frac{1.22 \lambda}{2 \text{NA}}, \quad (2.13)$$

where NA denotes the numerical aperture, which specifies the maximum angle to the optical axis for which the objective lens can accept in-focus light.

The Rayleigh limit is based on the visual criterion that two objects do not appear resolved if the image is not bi-modal. Therefore, it must not be confused with a strict limit of what can be concluded from an image. It may in fact be possible to precisely locate the two point sources, or to decide whether one or two sources are present in an image even if they are closer together than Δl . Such a decision can be greatly facilitated by prior knowledge, such as the relative brightnesses of the sources or the shape of the point spread function.

A more general quantification of the optical resolution is based on the point spread function itself. As stated in Section 2.1.2, the image lacks the high

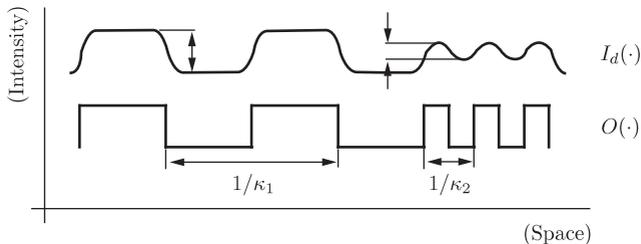


Figure 2.4: Blurring due to the point spread function reduces the modulation of high spatial frequencies.

spatial frequency information about the imaged fluorophore density (see Figure 2.4). Using the convolution theorem (Equation. 2.9), Equation 2.12 is reformulated as

$$I_d \propto \mathcal{F}^{-1}\{\mathcal{F}\{O\} \cdot \mathcal{F}\{P\}\} = \mathcal{F}^{-1}\{\mathcal{F}\{O\} \cdot \text{OTF}\}. \quad (2.14)$$

The *optical transfer function* OTF describes the spatial frequency response of the microscope. It is non-zero only over a finite range of spatial frequencies, which means that only the low-frequency part of $\mathcal{F}\{O\}$ can be observed. Beyond a certain cutoff frequency, all information is lost (see Figure 2.5). This view facilitates the understanding of what can be concluded from an ideal image in the best case.

In practice, *image noise* further obstructs the useful information in an image. Uncorrelated noise has a high-frequency spectrum and may therefore dominate regions in frequency space where the optical transfer function has small values.

2.1.4 DIGITAL IMAGES AND NOISE

In order to observe an image, the flux of photons needs to be integrated on some detector over a finite time. The intensity is then given by the photon count divided the integration time and area of the detector element. In order to achieve spatial sampling, the detector needs to be divided into small subunits. In the human eye, these subunits are the cone and rod cells

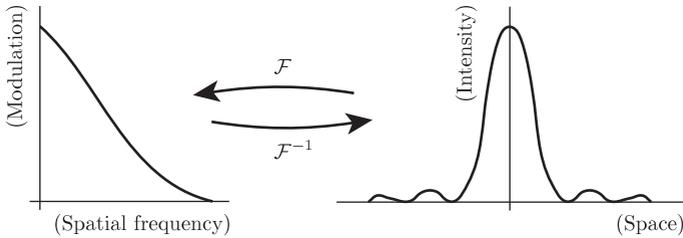


Figure 2.5: Relation between the modulation transfer function $\text{MTF} = |\text{OTF}|$ (left) and the point spread function (right).

that are arranged on the retina. In a digital microscope, these subunits are photo-sensitive electronic elements, for example charge-coupled devices (CCDs). Charge-coupled devices are essentially capacitors in which charges are separated by the photo-electric effect. They are typically arranged on a regular cartesian grid with constant spacing h . The individual detector elements are called *pixels*. A *digital image* is defined as a matrix \mathbf{I} with elements $I_{i,j}$, $(i, j) \in [1, N] \times [1, M]$ that correspond to the center locations $(x_i, y_j) = ((i - 0.5)h, (j - 0.5)h)$ of the pixels on the detector.

In practice, images are always corrupted by noise. Some of the noise arises from imperfect processing of the available information, while other components are due to fundamental physical limitations that can not be eliminated by clever engineering.

Photon emission is a *Poisson process*. That is, a constant intensity light source emits photons at a fixed rate, and the number of photons emitted per unit of time follows a Poisson distribution. The standard deviation of the number of emitted photons is equal to the square root of the mean. This type of noise is called *shot noise* and it can not be eliminated. Nevertheless, its detrimental effect decreases with increasing mean number of photons, which can be achieved by longer integration times or brighter light sources.

Further sources of noise include, but are not limited to, thermal fluctuations, imperfect readout of the CCD, and inaccurate conversion from an analog to a digital signal. Thermal fluctuations in the detector can cause

spurious photon detections. The charges in the CCD capacitors need to be read out and converted to a digital signal, which further increases the noise level. *Readout* and *conversion noise* are typically additive.

In electron-multiplying CCDs the number of charges is increased by an amplifier before they are further processed. This reduces the relative importance of readout and conversion noise. The amplification itself, however, is not perfect and introduces additional noise with a different distribution. This may lead to an overall improvement of the image quality for dim sources.

Due to the many different sources of noise, its distribution in a digital image is not obvious. Simple models include Gaussian additive noise, which does not depend on the signal (i.e. the expected number of photons), and Poisson noise, which does.

2.1.5 DECONVOLUTION

As shown in Section 2.1.2, fluorescence microscopy images are formed by a convolution of the fluorophore density with the point spread function resulting in a blurred image. *Deconvolution* [147, 63, 20] is a computational process that aims at reconstructing the true fluorophore density from the blurred image. Convolution is a linear operation and the discrete convolution can be seen as a linear transformation of the fluorophore density. Provided the point spread function is known, it should thus be possible to solve the system of linear equations of the transformation for the fluorophore density. It turns out, however, that this *inverse problem* is *ill-posed*, because the solution is not unique and highly sensitive to small perturbations in the image.

Figure 2.5 illustrates the relation between the frequency response of the microscope and its point spread function. The point spread function is band-limited. High spatial frequencies of the fluorophore density are partially suppressed and eventually lost. The solution of the inverse problem is, therefore, undetermined with respect to the spatial frequencies beyond the cutoff. Since frequencies beyond the cutoff do not pass the microscope, they can always be superimposed on any fluorophore density with-

out changing the image.

Reconstructing the true fluorophore density requires amplification of the high spatial frequencies in the image in order to counteract the damping caused by the optical transfer function. This, however, is highly sensitive to image noise, since image noise dominates the high frequency spectrum close to the cutoff. Without any means of *regularizing* the solution of the inverse problem, deconvolution thus fails.

In the following sections I will briefly review some common approaches to regularized deconvolution.

2.1.5.1 Wiener Deconvolution

The idea of *Wiener filtering* is to only amplify spatial frequencies for which the signal is not dominated by noise. The fluorophore density is reconstructed in frequency space as [122]:

$$\mathcal{F}\{O\} \approx \mathcal{F}\{I_d\} \cdot \frac{1}{\text{OTF}} \cdot \frac{|\text{OTF}|^2}{|\text{OTF}|^2 + \frac{N}{S}}, \quad (2.15)$$

where N and S are the mean power spectral densities of the noise and the signal (i.e., of $O(\cdot)$), respectively. Because N/S is non-zero everywhere, the denominator of Equation 2.15 remains finite. Since S is related to the unknown quantity $O(\cdot)$, it has to be estimated based on prior knowledge. Therefore, N/S allows to conveniently quantify which part of the frequency spectrum is significant – or trustworthy. Note that the Wiener filter implicitly assumes additive Gaussian image noise.

2.1.5.2 Tikhonov Deconvolution

By discretizing both sides of Equation 2.12, one obtains the fluorescence microscopy imaging model in matrix–vector form

$$\mathbf{i}_d = \mathbf{P}\mathbf{o}, \quad (2.16)$$

where \mathbf{i}_d and \mathbf{o} are the sampled values of the image and fluorophore density arranged in vectors, and \mathbf{P} is the blurring matrix constructed from the point spread function. *Tikhonov filtering* (or more specifically *Tikhonov deconvolution*) minimizes the functional

$$\Phi(\hat{\mathbf{o}}) = \|\mathbf{P}\hat{\mathbf{o}} - \mathbf{i}_d\|^2 + \lambda\|\mathbf{C}\hat{\mathbf{o}}\|^2, \quad (2.17)$$

where the second term, weighted by the regularization parameter λ (not to be confused with the wavelength), penalizes high frequency oscillations dominated by noise. Minimization of Equation 2.17 yields [122]:

$$\hat{\mathbf{o}} = (\mathbf{P}^T\mathbf{P} + \lambda\mathbf{C}^T\mathbf{C})^{-1}\mathbf{P}^T\mathbf{i}_d. \quad (2.18)$$

Tikhonov and Wiener deconvolution (Section 2.1.5.1) are closely related and can be made identical by appropriate choice of \mathbf{C} . While in Tikhonov deconvolution the regularization happens in real space, Wiener deconvolution constrains the solutions in frequency space. Both methods are very sensitive to the correct specification of the point spread function and do not ensure positivity of the solution $O(\cdot)$ [122].

2.1.5.3 Iterative Deconvolution

Wiener and Tikhonov filtering, and similar approaches to deconvolution, find solutions in a single computational step (see Equations 2.15 and 2.18). In case of non-trivial constraints on the solution (such as positivity) or non-Gaussian noise models, it may only be possible to construct iterative algorithms (see [122] for a list of various methods).

Notable conceptual extensions of the two methods discussed so far are statistical approaches to deconvolution, such as *maximum likelihood* or *maximum a posteriori deconvolution*. The idea is to find the $O(\cdot)$ that is in best agreement with the data under the given noise model, while fulfilling the constraints given by the imaging model (e.g. Equation 2.12, positivity of the solution, etc.) In case of the maximum a posteriori estimator, “best agreement” also refers to an additional prior on $O(\cdot)$.

A great advantage of these methods is that they yield results that can be interpreted in a statistical sense. Furthermore, they allow integrating prior knowledge in more intuitive ways than, for example, the specification of the term N/S or the matrix \mathbf{C} in Wiener and Tikhonov deconvolution, respectively.

2.1.5.4 2-Dimensional versus 3-Dimensional Deconvolution

In wide-field microscopy, a considerable amount of the blurring can be attributed to out-of-focus light. Theoretically, this is not different from lateral blurring within one focal plane, and it is fully defined by a 3D point spread function. Nevertheless, many deconvolution methods treat the axial and radial dimensions of 3D images differently. 3D deconvolution methods require stacks of images that are acquired by sequentially focussing in different planes along the optical axis.

A simple approach to removing out-of-focus light is to subtract blurred versions of the images in adjacent planes from each image in the axial stack of images. *Nearest neighbor methods* only consider the planes immediately above and below the current plane of interest. *Multi-neighbor methods* consider more than one neighbor above and below. *No neighbor methods* only consider blurring within, but not between planes. They are hence purely 2-dimensional. Their application is limited to objects with high spatial frequencies or effectively flat samples [100]. In confocal microscopy, axial resolution (and to a lesser extent lateral resolution) is greatly enhanced, and 3-dimensional deconvolution may not be necessary.

2.1.6 EMPIRICAL CHARACTERIZATION OF MICROSCOPES

Real imaging systems are hardly ideal, as reflected by more or less severe imperfections. These imperfections include point spread functions with non-zero widths, *chromatic aberration*, and lateral variations in the illumination. Simple protocols for determining the point spread function (Section 2.1.6.1), *chromatic aberration* (Section 2.1.6.2), and lateral variations in the illumination (Section 2.1.6.3) are hence presented here.

2.1.6.1 Determining the Point Spread Function

In Section 2.1.5 it was discussed how to improve image quality by deconvolution. Except for *blind deconvolution* (which was not discussed), all deconvolution methods require knowledge about the point spread function. Some methods are even highly sensitive to its precise shape. The discrepancy between a theoretical and a measured point spread function can be quite significant in practice. Whenever possible, empirical point spread functions should thus be used. A theoretical point spread function defined in terms of Bessel functions is given in References [89, 47, 122, 61].

Empirical point spread functions can be estimated from images of point-like fluorescent objects. “Point-like” means that the diameter should be significantly smaller (say, by a factor of 4) than the typical width of the model point spread function. For a confocal microscope with $NA = 1.4$ and emission wavelength $\lambda = 532$ nm, the full width at half maximum of the model point spread function is approximately 461 nm. Therefore, 100 nm fluorescent beads could be used.

Images of such beads (or any other small object) should be acquired with a high signal-to-noise ratio and fine spatial sampling. If a 3-dimensional point spread function is needed, image stacks need to be recorded. First, the (3-dimensional) location of the bead needs to be determined, e.g. by estimating the intensity centroid [125]. Assuming radial symmetry, the value of the point spread function is then determined by averaging interpolated intensities along concentric circles centered at the intensity centroid [131]. Moving the circles in axial direction allows determining a 3-dimensional point spread function.

2.1.6.2 Quantifying Chromatic Aberration

Chromatic aberration is wavelength-dependent spatial distortion of the image. The microscope fails to focus different colors originating from the same source onto the same location in the image. This may corrupt measurements of spatial correlations between objects imaged in different colors.

The magnitude of chromatic aberrations typically depends on the location in the image. The lateral shift between the focal points in the image plane often increases with increasing distance from the optical axis (i.e., the center of the image). A simple approach to visualizing the aberration as a function of location is to image small point sources, such as multi-color fluorescent beads, in different color channels. The distribution of beads needs to cover the whole field of view, but individual beads should not touch. The bead locations can be determined by intensity centroid estimation [125], and subsequently compared across the different color channels. This yields a map of *relative* distortions for each pair of colors. If the true bead location can be determined by other means, *absolute* distortion maps can be computed.

In most cases, the (lateral) shifts Δx and Δy between colors depend linearly on the x - and y -positions of the beads [78]. This allows conveniently estimating a calibration function by two linear least squares fits to observed pairs $\{(x_i, \Delta x_i)\}$ and $\{(y_i, \Delta y_i)\}$. This calibration function can later be used to correct location estimates of other imaged structures.

2.1.6.3 Lateral Variation in Illumination

For homogeneous illumination of the sample, the observed fluorescence intensity is roughly proportional to the fluorophore density. This allows creating maps of protein localization in cells in a straightforward way. In real imaging systems, however, the illumination is not, or only approximately, homogeneous. The effect of inhomogeneous illumination can be corrected as follows: Prior to imaging the sample of interest, the illumination pattern can be estimated from a calibration sample with spatially homogeneous fluorophore density. Images of the real sample can then be normalized with the image of the calibration sample.

2.2 SEGMENTATION OF INTRA-CELLULAR STRUCTURES

Two approaches to extracting shape information from images can be distinguished: Image-model-based and object-model-based methods.

Image-model-based methods extract information directly from the images. In the simplest case, image transformations are applied to highlight features of interest such as edges or peaks of intensity. Edges can also be used to segment the image into fore- and background. A complementary approach to *image segmentation* (see e.g. [115] and references therein) is to fit a model for the statistics of pixel intensities of the fore- and background, such as the mean intensity within segments, their texture, and so on. Applications of these methods range from the anatomical to the molecular scale [39]. A classical problem is the segmentation of images of cells [141, 38, 93, 29, 135].

The term “image model based” can be understood in the sense that these methods describe the structure of the data on the level of the images, while neglecting the image formation process. Therefore, image-model-based methods rely on the implicit assumption of a one-to-one correspondence between the features of the image and the true object.

Some authors have pointed out that if the size of the imaged structures is comparable to the width of the point spread function of the imaging device, standard image segmentation can not be considered an unbiased estimate of the object’s shape [141, 38]. That is, the assumption of a one-to-one correspondence between image features and features of the true object breaks down. In fact, many intra-cellular structures have sizes on the order of 10 to 1000 nm, whereas the full width at half maximum (FWHM) of standard microscopy techniques is in the range of 200 to 500 nm.

While object shapes are obscured by the blurring due to the point spread function, object symmetries are preserved. For point-symmetric objects smaller than the FWHM, the observed intensity centroid hence coincides with the true one. Besides this centroid, however, not much information can be reconstructed. A similar argument holds for line-like objects, whose true width needs to be neglected. In these two cases, fitting simple geometric shape models to image features provides good approximations to the true objects’ locations [23, 125, 87].

For objects that are approximately the size of the point spread function or larger, outline shape information can be reconstructed if an image model that accounts for the blurring due to the imaging process is fit to the observed image. The resulting **object-model-based** approach tries to

explain the image in a bottom-up way using (physical) object and imaging models. The structure of the data is thus explored on the level of the imaged objects, rather than the image itself.

Reconstructing physical object boundaries from blurred microscopy images is equivalent to solving a combination of a segmentation and a deconvolution problem. Deconvolution is known to be ill-posed [131, 122], and direct linear deconvolution is generally not recommended for sub-cellular structures [103]. Constrained (or regularized) iterative deconvolution methods allow enforcing specific behavior of the solution. However, they simply produce a de-blurred image, leaving the segmentation problem unsolved.

The approach I follow here is to implicitly regularize the deconvolution problem by the object and imaging models that are used for the solution of the segmentation problem. The deconvolution is, therefore, just a side-product [103]. The present approach requires prior knowledge about the shape of the imaged objects. Such prior knowledge increases the detection power and regularizes the solution, but it decreases the exploration power. Here, the prior is encoded in restrictions on the geometrical model for the imaged objects. Objects are described by their *outlines* and the fluorophore density inside the outlines. The outlines are assumed to be smooth on at least the length scale corresponding to the cutoff frequency of the microscope, but may also be smooth on larger scales.

Explicit active contours are a widely used image segmentation method that allows enforcing regularity (smoothness) [72] and topology [30] of segment outlines. Outlines are parameterized by (piece-wise linear) splines in a compact way. I use an adaptation of the classical active contour formulation to solve an iterative constrained deconvolution problem. The method is well-suited to segment small, compact, intra-cellular structures, such as endosomes or the Golgi. The method is tested on a variety of synthetic data (Section 2.2.3). An extensive study of the morphology of endosomes (Section 2.2.4.3 and 2.2.4.4) and their co-localization with virus particles (Section 2.2.4.5), as well as a small demonstration on images of the Golgi (Section 2.2.4.1), are further presented.

2.2.1 DEFINITIONS AND PROBLEM STATEMENT

The outline of an object is a curve that approximately describes the intersection of the physical boundary of the real object and the imaged plane. The *outline* of the k -th object in an image is represented by a piece-wise linear closed spline through the n_k *support points* $\mathbf{v} = (x, y)^T \in \mathbb{R}^2$

$$\Theta^k = \{\mathbf{v}_1^k, \dots, \mathbf{v}_{n_k}^k\}, \quad (2.19)$$

where $\mathbf{v}_{n_k}^k = \mathbf{v}_1^k$. The set $S = \{\Theta^k\}_{k=1}^N$ holds a collection of N objects, described by their outlines Θ .

The goal of *intra-cellular object outline reconstruction* is to find a parametric description of a set S of outlines of objects, supported by an error-corrupted digital image $\mathbf{I}_m = \mathbf{I} + \varepsilon$. Given an imaging model $\mathbf{I}(S)$, the parameters Θ of the set S of object outlines have to be found that best explain the measured image \mathbf{I}_m , while keeping the complexity of the outlines minimal. The imaging model $\mathbf{I}(S)$ predicts the image \mathbf{I} of the set S of objects in the absence of noise.

2.2.2 ALGORITHM

The above parameter estimation problem can be rephrased in the context of classical explicit active contours. The idea of active contours is to evolve a parametric curve toward “interesting” features in the image, while ensuring regularity of the curve through internal soft constraints. In the following sections I describe how to extend the original explicit active contours framework to *deconvolving active contours*, and how to use them to accurately segment small intra-cellular objects from noisy fluorescence microscopy images.

2.2.2.1 Active Contours

The original idea of active contours is to match a deformable model to interesting image features by energy minimization. The choice of the energy

functional thus implicitly defines the image analysis task to be solved. The energy is typically divided into two terms:

$$E(\mathbf{I}, \Theta) = E_{\text{int}}(\Theta) + E_{\text{im}}(\mathbf{I}, \Theta), \quad (2.20)$$

namely an *internal energy* E_{int} of the contour Θ , which ensures regularity (smoothness), and an *image energy* E_{im} , which is locally minimal on prominent features of the image \mathbf{I} , such as edges. The internal energy consists of terms E_b and E_s for bending and stretching of the contour, respectively:

$$\begin{aligned} E_b(\Theta) &= \beta \|\mathbf{v}_{n-1} - 2\mathbf{v}_1 + \mathbf{v}_2\|^2 \\ &\quad + \beta \sum_{i=2}^{n-1} \|\mathbf{v}_{i-1} - 2\mathbf{v}_i + \mathbf{v}_{i+1}\|^2 \quad \text{and} \\ E_s(\Theta) &= \alpha \sum_{i=2}^n \|\mathbf{v}_{i-1} - \mathbf{v}_i\|^2, \end{aligned} \quad (2.21)$$

where the fact that the first and last point of the outline are identical was used. The parameters α and β allow adjusting the relative weight of bending, stretching, and image energy. Minimizing the energy functional in Equation 2.20 gives rise to two independent (for x and y) Euler equations [72, 25]. In discretized vector form, the Euler equations for inner points of the outline are:

$$\begin{aligned} &\alpha (-\mathbf{v}_{i-1} + 2\mathbf{v}_i - \mathbf{v}_{i+1}) \\ &+ \beta (\mathbf{v}_{i-2} - 4\mathbf{v}_{i-1} + 6\mathbf{v}_i - 4\mathbf{v}_{i+1} + \mathbf{v}_{i+2}) \\ &+ (f_x(i), f_y(i))^T = 0, \end{aligned} \quad (2.22)$$

where $f_x(i) = \partial E_{\text{ext}}/\partial x_i$ and $f_y(i) = \partial E_{\text{ext}}/\partial y_i$ denote the partial derivatives of the external energy with respect to the location of the i -th support point $\mathbf{v}_i = (x_i, y_i)^T$. For the points v_i with $i \in \{1, 2, 3, n-1, n\}$ the periodicity of the outline is used to adapt the equations accordingly. The Equations 2.22 can be written in compact matrix form as:

$$\begin{aligned} \mathbf{A}\mathbf{x} + \mathbf{f}_x(\mathbf{x}, \mathbf{y}) &= \mathbf{0} \\ \mathbf{A}\mathbf{y} + \mathbf{f}_y(\mathbf{x}, \mathbf{y}) &= \mathbf{0}. \end{aligned} \quad (2.23)$$

\mathbf{A} is a pentadiagonal banded matrix that defines the forces acting on the support points of the outline as a response to local bending and stretching. Equation 2.23 thus describes a state in which internal and external forces (as given by the gradient of the image energy) are balanced. An outline satisfying Equation 2.23 can be found [72] by iterating

$$\begin{aligned}\mathbf{x}_t &= (\mathbf{A} + \gamma\mathbf{1})^{-1} (\mathbf{x}_{t-1} - \mathbf{f}_x(\mathbf{x}_{t-1}, \mathbf{y}_{t-1})) \\ \mathbf{y}_t &= (\mathbf{A} + \gamma\mathbf{1})^{-1} (\mathbf{y}_{t-1} - \mathbf{f}_y(\mathbf{x}_{t-1}, \mathbf{y}_{t-1}))\end{aligned}\tag{2.24}$$

until convergence, where γ is the step size, $\mathbf{1}$ the identity matrix, and t a pseudo time.

In the original explicit active contour formulation (not the present adaptation), $f_x(i)$ and $f_y(i)$ only depend on local image features and can hence be precomputed. $(\mathbf{A} + \gamma\mathbf{1})^{-1}$ can also be precomputed, since it does not depend on the current outline. Equation 2.24 is implicit with respect to the internal forces and explicit with respect to the external forces. As long as the external forces are not too large, large step sizes can be used. For these reasons, the algorithm in Equation 2.24 rapidly converges to a local minimum of Equation 2.20.

In Equation 2.24 the external forces appear in only one term, which is independent from the internal forces. This allows to easily integrate different models. In fact, Equation 2.24 can be seen as a way to regularize any task of fitting outlines to data. In Section 2.2.2.2 an image energy is constructed that is small when the outline coincides with the true outline of an object imaged with a fluorescence microscope. In Section 2.2.2.4 it is shown how to efficiently compute the gradient of this image energy.

2.2.2.2 Image Energy

The image energy is based on the similarity between the observed image \mathbf{I}_m and a model image for a set S of outlines. For now, however, I will restrict the specification of the model to the case where only one object is present in the image, that is $S = \{\Theta^1\}$. As long as objects are further apart than the width of the point spread function, the outline refinement problem for

multiple objects in one image can be reduced to independent instances of the single object case. The case of multiple close objects requires a few extensions that will be discussed in Section 2.2.2.6.

Object outlines enclose regions of high fluorophore density. The fluorophore density $O(\cdot)$ is discretized on a grid with spacing h , that is, $O_{i,j}^h = O(ih, jh)$. Given an outline Θ , the elements $O_{i,j}^h$ of the matrix \mathbf{O}^h are defined as:

$$O_{i,j}^h = \begin{cases} c & \text{if } (ih, jh) \text{ enclosed by } \Theta \\ (1-d)c & \text{if } d = D((ih, jh), \Theta) < 1 \\ 0 & \text{else} \end{cases}, \quad (2.25)$$

where D is the distance to the outline Θ and c the constant fluorophore density inside the object. Similar to the simplified Mumford-Shah functional [22], this object intensity function is piecewise constant, but with linearly decaying intensities at the boundaries, as illustrated in Figure 2.6. The linear decay makes \mathbf{O}^h continuous with respect to the outline Θ , which will be important for stability and convergence of the energy minimization algorithm (Equation 2.24).

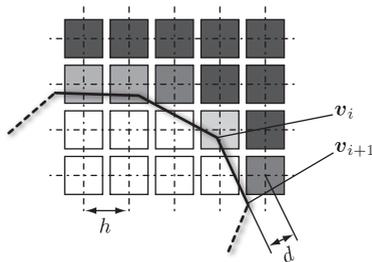


Figure 2.6: Construction of the fluorophore density \mathbf{O}^h from an outline. Bright squares represent large values of $O_{i,j}^h$.

The piecewise linear fluorophore density is favored over more complex models as it requires estimating only a minimum number of parameters. On noisy data this increases the robustness of parameter estimation. Robustness is the central objective if no prior knowledge about the true fluorophore density is available. Robustness of parameter estimation is directly

related to the regularization of the deconvolution. Similar to Tikhonov deconvolution (see Section 2.1.5.2), the solution is strongly constrained in real space, since variations of \mathbf{O} are limited by its implicit definition through the outline Θ .

Image formation in fluorescence microscopy is modeled by a convolution of the fluorophore density with the point spread function (see Section 2.1.2, in particular Equation 2.12). In discretized form, the image model is:

$$\mathbf{I}^h(\Theta) = \mathbf{O}^h * \mathbf{P}^h, \quad (2.26)$$

where $*$ now denotes the discrete convolution operator. The convolution kernel \mathbf{P}^h is the point spread function of the microscope, discretized with spacing h . \mathbf{O}^h is implicitly defined by the outline Θ through Equation 2.25. The sample points (ih, jh) can coincide with the pixel centroids of the measured image \mathbf{I}_m . In practice, however, finer resolutions h can be used, which may be beneficial for fitting finer structures in the data. In order to match the resolution of \mathbf{I}_m , finely sampled images \mathbf{I}^h have to be down-sampled.

The image energy associated with outline Θ is given by the weighted squared error between the pixel intensities $(\mathbf{I}_m)_{i,j}$ and $I_{i,j}$ of the observed and model images (assuming the resolutions h match):

$$E_{\text{im}} = \sum_i \sum_j R_{i,j} \left((\mathbf{I}_m)_{i,j} - I_{i,j} \right)^2. \quad (2.27)$$

The elements $R_{i,j}$ of the weighting matrix \mathbf{R} reflect the trustworthiness of the measurements \mathbf{I}_m . If all $R_{i,j} = 1$ the image energy is the squared error between model and observation. Using this error measure amounts to (implicitly) assuming normally distributed noise with identical variance for all pixels. As discussed in Section 2.1.4, noise in images is a superposition of different sources. The fundamental limit is given by the shot noise, which imposes a Poisson distribution for the number of detected photons. Due to the conversion of units from number of photons to intensities, however, the Poisson property is lost. That is, one can not simply assume that intensities are Poisson-distributed. Additional sources of noise further complicate the situation, since they may introduce additional normally distributed noise, or even noise with an unknown distribution. Unless

detailed knowledge about the distribution of pixel intensities is available, a conservative choice is to use the squared error. According to the *maximum entropy principle* [69], the distribution that best represents the available knowledge is the one with the highest entropy. If only the mean and the variance of the noise are known, the normal distribution is the one with highest entropy, thus justifying the use of the squared error as a measure of similarity between model and data.

Equations 2.20, 2.25, 2.26, and 2.27 form the *deconvolution functional* for “deconvolving explicit active contours”.

2.2.2.3 Determination of the Fluorophore Density

In the previous section the deconvolution functional was introduced. It involves mapping Θ to \mathbf{O} , then \mathbf{I} , and finally to E_{im} . Up to the multiplicative constant c , this mapping is fully defined.

In principle, c could be estimated along with Θ , for instance, by alternating minimization steps of the deconvolution functional with respect to Θ and c . In practice, however, the constant c is better estimated a priori.

Estimation of c is based on the observation that the central intensity of imaged objects is related to their true fluorophore density and to their size. For objects that are far larger than the width of the point spread function, the fluorophore density can be identified with the intensity ϕ in the center of the image of the object (right image in Figure 2.7). Smaller objects have a central intensity that is reduced by a factor κ .

In order to estimate κ for a given outline Θ , the radial intensity profile $J(r)$ of the object (Figure 2.7), found by averaging interpolated intensities along concentric circles around the intensity centroid, is analyzed. This is the same procedure as the point spread function estimation described in Section 2.1.6.1, with the difference that now the imaged objects are not necessarily point sources.

The half width at half maximum (HWHM) $r_{0.5}$ of $J(r)$ serves as a size parameter (Figure 2.7). Since the dependence $\kappa = f(r_{0.5})$ is not explicitly known, it is empirically calibrate on synthetic images. These images \mathbf{I}^i

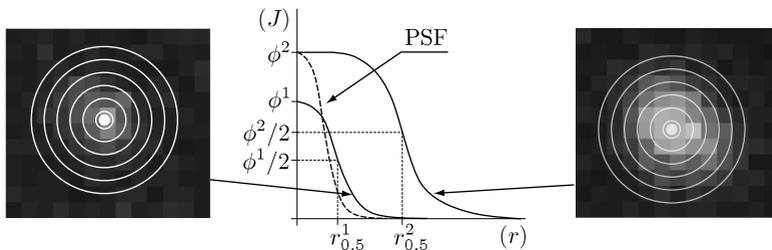


Figure 2.7: Radial intensity distribution of different objects as found by averaging intensities along concentric circles around the intensity centroid. As objects get smaller, the radial intensity distribution develops a peak and the central intensity is reduced.

are generated by convolving (Equation 2.26) fluorophore densities (Equation 2.25) with the point spread function. The sizes of the objects are varied and the constant fluorophore density c^i remains fixed. For each synthetic image I^i the size parameter $r_{0.5}^i$ and the central intensity ϕ^i is measured. The calibration function is then computed as $\bar{\kappa}^i(r_{0.5}^i) = c^i/\phi^i$. Based on this function, the object intensity c^k of an experimentally observed object k can later be estimated as $c = \phi\kappa(r_{0.5})$ using (linear) interpolation. The calibration function $\kappa(r_{0.5})$ is valid for all objects, regardless of their true intensity.

For very small, point-like objects, $r_{0.5}$ converges to the HWHM of the point spread function and does no longer vary with object size. Hence, $\kappa(r_{0.5})$ is asymptotically approaching infinity at the HWHM of the point spread function and estimates of c degenerate. Without a robust estimate of c , analysis of shapes and sizes of objects is impossible. As discussed above, estimating c fails when the imaged object is too small. This effect can be understood in the theory of Fourier optics (Section 2.1.2): The imaged object can be represented by a power spectrum of spatial frequencies. The higher spatial frequencies of this spectrum are needed to define the location of the object boundary. Since the microscope behaves like a low-pass filter in spatial frequency space, size information is lost. For a very small object, variations of the size or fluorophore density become indistinguishable in the image. Hence, these parameters can not be individually determined.

The analysis of sizes and shapes is, therefore, restricted to objects with $r_{0.5}$ above an empirically determined threshold. Depending on the image quality, this threshold is found to be roughly 1.1–1.5 HWHM. Smaller objects are treated as circles with centers at the observed intensity centroid. Their radius is chosen as the minimum radius of the calibration objects for which $r_{0.5}$ was above the threshold. Fluorophore densities c can be found by least squares regression on Equations 2.27 and 2.26.

The procedure for determining c relies on the validity of the empirical calibration function $\kappa(r_{0.5})$. Since circular objects are used for calibration, one must expect a bias in c for objects that are non-circular in reality. As will be shown in Section 2.2.3, the bias is acceptable for objects that have a globular, rather than elongated or highly concave, shape.

2.2.2.4 Gradient of the Image Energy

In the classical explicit active contour formulation the image energy – and hence its gradient – only depends on the location of individual support points in the image. This allows using a lookup table with precomputed values of the image energy gradient.

Here, the image energy of the deconvolution functional has a different structure with respect to the support point locations. As illustrated in Figure 2.6, outlines determine the fluorophore density \mathbf{O}^h , which is the basis of the image energy. The value of the fluorophore density of the lower-right pixel in Figure 2.6, for example, depends on the location of the support points \mathbf{v}_i and \mathbf{v}_{i+1} . Furthermore, the image energy is not linear in the elements of \mathbf{I} , which, according to the image formation model (Equation 2.26), depend on several elements of \mathbf{O}^h . Taken together, the energetic effect of changing one support point location depends on the local shape of the outline in a non-trivial way. This requires approximating the partial derivatives of E_{im} with respect to the positions $\mathbf{v}_i = (x_i, y_i)^T$ of all support points i in each iteration. I use a finite-difference approximation to the derivatives with respect to the positions of the support points \mathbf{v}_i of an outline Θ .

For a given outline Θ the image energy E_{im} is computed as described

(Equation 2.27). In order to compute the finite difference approximations to the gradient, support points \mathbf{v}_i are moved a small distance s in direction of the local outline normal $\mathbf{n}_i = ((n_x)_i, (n_y)_i)^T$, $|\mathbf{n}_i| = 1$ (see Figure 2.8). Moving the support point parallel to the outline does not change the image energy to a first approximation. The move in normal direction yields a deformed outline Θ_* . The corresponding fluorophore density \mathbf{O}_*^h , model image \mathbf{I}_*^h , and new image energy $E_{\text{im},*}$ are then computed, and finally:

$$\frac{\partial E_{\text{im}}}{\partial \mathbf{n}_i} \approx \frac{E_{\text{im},*} - E_{\text{im}}}{s}. \quad (2.28)$$

The derivatives with respect to x_i and y_i follow by projection:

$$\frac{\partial E_{\text{im}}}{\partial x_i} \approx (n_x)_i \frac{\partial E_{\text{im}}}{\partial \mathbf{n}_i} \quad (2.29)$$

and

$$\frac{\partial E_{\text{im}}}{\partial y_i} \approx (n_y)_i \frac{\partial E_{\text{im}}}{\partial \mathbf{n}_i}. \quad (2.30)$$

The computation of E_{im} is costly. The cost is high for mainly three reasons: First, constructing \mathbf{O}^h requires deciding whether a point $(ih, jh)^T$ is inside the outline or not, which depends on all support points of the outline. Second, the distance d to the closest point on the outline needs to be determined for all locations $(ih, jh)^T$ close to the outline. Third, the discrete convolution of \mathbf{O}^h with the point spread function (Equation 2.26) needs to be computed, which scales as $N^2 \times M^2$ for a fluorophore density \mathbf{O}^h and point spread function \mathbf{P}^h with sizes N^2 and M^2 , respectively. However, E_{im} needs to be computed only once in each iteration of the energy functional minimization. In contrast, $E_{\text{im},*}$ needs to be computed for *each support point* in each iteration. An efficient algorithm to determine $E_{\text{im},*}$ is therefore highly desirable.

Since convolution is a linear operation, the model image of the deformed outline can be expressed as

$$\mathbf{I}_*^h = \mathbf{I}^h + \Delta \mathbf{I}_*^h, \quad (2.31)$$

where $\Delta \mathbf{I}_*^h$ denotes the change in \mathbf{I}^h caused by the deformation of Θ into Θ_* . This change is found by computing \mathbf{O}_*^h from Θ_* , subtracting it from

\mathbf{O}^h , and convolving this difference $\Delta \mathbf{O}_*^h$ with the discretized point spread function \mathbf{P}^h :

$$\Delta \mathbf{I}_*^h = \Delta \mathbf{O}_*^h * \mathbf{P}^h. \quad (2.32)$$

As illustrated in Figure 2.8 the deformation of a single support point only influences a few elements of \mathbf{O}^h . The computational cost can therefore be drastically reduced by only considering locations $(ih, jh)^T$ in a small region around the support points \mathbf{v}_{i-1} , \mathbf{v}_i , and \mathbf{v}_{i+1} . This region is given by all locations contained in a curve that is constructed by offsetting the convex hull of the original and displaced support points \mathbf{v}_{i-1} , \mathbf{v}_i , and \mathbf{v}_{i+1} outward by $1h$. The cost of computing \mathbf{I}_*^h thus scales as $K \times M^2$, where K is the number of elements in \mathbf{O}^h that are influenced by moving a single support point of Θ . This number is proportional to the length of the two outline segments adjacent to the moved point \mathbf{v}_i (dotted lines in Figure 2.8). The total cost of computing a $\Delta \mathbf{I}_*^h$ for all support points is therefore roughly proportional to the length of the outline, regardless of the number of support points used. Computation of each $E_{\text{im},*}$ scales as N^2 . This cost can be further reduced by only recomputing summands $R_{i,j}((I_{\text{m}})_{i,j} - (I_*)_{i,j})^2$ in Equation 2.27 that are changed by the deformation of the outline.

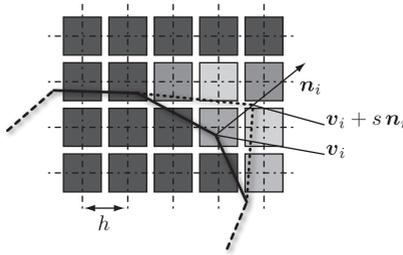


Figure 2.8: Changes of \mathbf{O}^h upon changing support point locations are local and continuous. Bright squares represent large changes in $\mathbf{O}_{i,j}^h$.

2.2.2.5 Minimizing the Energy Functional

Given the gradient of the image energy with respect to the support point locations, Equation 2.24 is used to update the current outline. The step

size parameter γ is typically a compromise between stability and speed of convergence. Since the update rule is implicit with respect to the internal forces, but explicit with respect to the external forces, an upper limit for γ depends on the magnitude of the largest image energy gradient.

Here, the steps of the algorithm are not directly controlled by γ , but rather by the sampling distance s used for computing the finite difference approximation of the gradient (Equation 2.28). As long as the displacements of the support points lower the image energy, the algorithm is stable with respect to the external forces.

The idea is to bound the support point displacements caused by the image energy gradient. Ignoring for a moment the internal forces, no support point should move further than s . This is achieved by adjusting γ such that the support point i with the largest image energy gradient makes a step that brings it approximately to $\mathbf{v}_i + s \mathbf{n}_i$. Since the effective step made is proportional to the image energy gradient, all other support points will make steps smaller than s . No support point will hence move across a local image energy minimum.

This strategy, however, does not take into account correlations between the image energy gradients of neighboring support points. As a consequence, oscillations of the outline may appear. Furthermore, it is desirable to *control* the effective step size, rather than only *limiting* it. A tradeoff between speed of convergence and oscillations is found by a simple heuristic for adapting s : Between application-specific upper and lower bounds, s is increased by a constant factor whenever the image energy was reduced in two consecutive iterations, and decreased by the same factor otherwise. This update rule yields a minimization algorithm that quickly moves the outline over regions of small image energy gradients, but allows for sufficiently fine resolution close to a local minimum.

The stopping criterion of the energy minimization is based on the change in the outline. This follows the logic that the outline should minimize the image energy subject to a smoothness constraint, rather than the image energy alone. The change of the outline is quantified by the mean change of the support point locations $\langle \Delta \mathbf{v}_i \rangle$ in an iteration. The minimization algorithm is stopped when this value drops below an application-specific threshold.

2.2.2.6 Multiple Objects

So far it was assumed that only a single object is present in the image, or that objects are well-separated such that their respective images do not interfere. Due to the blurring caused by the point spread function, images appear wider than the region with non-zero fluorophore density of the underlying objects. If two objects are closer than the width w (say, three times the half width at half maximum) of the point spread function, some pixels in the image are influenced by both objects (see Figure 2.9A). Due to the superposition principle (afforded by the linearity of the image formation model in Equation 2.26), the images add up. The image energy of one outline is thus no longer independent of the other.

The outlines of groups S of objects with coupled image energies need to be estimated simultaneously. The algorithm is essentially the same, except that the reference image energy E_{im} in Equation 2.28 needs to be computed on the basis of the joint fluorophore density of the group of objects. This means that \mathbf{O}^h is created for a group of outlines, convolved with the point spread function, down-sampled (if needed), and then used in Equation 2.27 to get the image energy. The image energy gradient is computed with respect to all support points of all outlines before the outlines are simultaneously updated.

The objects are grouped according to the location of initial outlines (initialization will be discussed in Section 2.2.2.7). Any two (or more) objects that influence each other must be assigned to the same group. A solution to this problem can be found by testing all close objects for mutual influence and grouping them together if required. If a tested object is already assigned to another group, groups are merged. For each group a region of interest is found that includes all pixels in the image influenced by any object in the group. This yields a *decomposition* of the image \mathbf{I}_m into (not necessarily disjoint) *sub-images* \mathbf{I}^i (Figure 2.9B), which, together with the corresponding group S^i of objects, can be processed independently of the others.

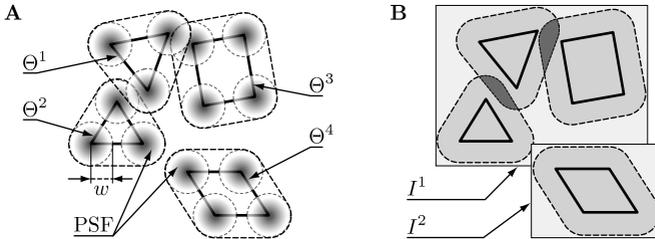


Figure 2.9: Grouping of coupled objects (A) and decomposition of the image into sub-images (B) that encompass all pixels influenced by the associated group. I^1 is associated with $S^1 = \{\Theta^1, \Theta^2, \Theta^3\}$, and I^2 with $S^2 = \{\Theta^4\}$

2.2.2.7 Initialization

Minimization of the energy functional (Equation 2.20) with the algorithm of Equation 2.24 requires an initial estimate of the outlines. The initial outlines are important for two reasons: First, they fix the number of objects in the image a priori, which further regularizes the deconvolution by drastically reducing the space of possible joint fluorophore densities (that is, the fluorophore density of a collection of objects). Second, they constrain the convergence of the energy functional minimization algorithm to a nearby set of local minima. Initial outlines can be found by standard image processing algorithms. In the following, I outline an initialization recipe that is equally applicable to the benchmark examples in Section 2.2.3 and the case study considering images of endosomes in Section 2.2.4. It consists of background subtraction, Gaussian filtering and edge highlighting, and watershed segmentation. Other data may require different recipes or parameter values.

In a first step, low-frequency background variations have to be removed from the image since they comprise all non-object-specific signals that can not be accounted for by the object model (Equation 2.25). I use a method related to the rolling ball algorithm [139]: For each pixel in the unprocessed image \mathbf{I}_m , the local background value is determined as the most frequently occurring intensity value in square region (typically 15×15

pixels) centered at that pixel. The so-found background image is then smoothed with a Gaussian kernel ($9 \times 9, \sigma = 2$) to reduce local variations (see Figure 2.10, middle). Subtracting the smoothed background image from the input image (Figure 2.10, left) yields a virtually background-free image (Figure 2.10, right).

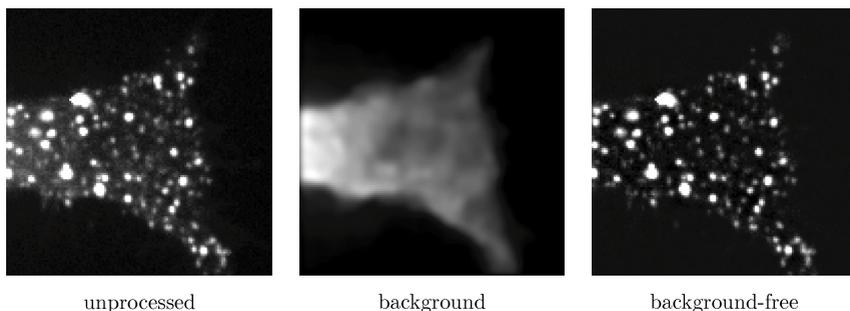


Figure 2.10: Example of background removal from an image of Rab5-EGFP endosomes in a part of an human embryonic retinoblast (HER) 911 cell. Images are contrast enhanced (but *not* thresholded).

In the next step, objects are detected in the image by means of image segmentation. An initial outline is given by the closed path of pixels around a foreground segment. Images of small globular objects typically contain one dominant local intensity maximum per object. Each object is therefore enclosed by a ridge in the spatial gradient or Laplacian of the intensity. These intensity edges are highlighted by thresholding each pre-processed image followed by a convolution with a Laplacian-of-Gaussian kernel ($5 \times 5, \sigma = 0.2$). The result is shown in Figure 2.11, left panel. The watershed transform [149] is then applied to detect pixels that lie on ridges of the edge image and are, hence, close to the outline of a globular object (Figure 2.11, right). Connecting neighboring ridge pixels to closed paths yields the desired initial outlines to single-pixel resolution.

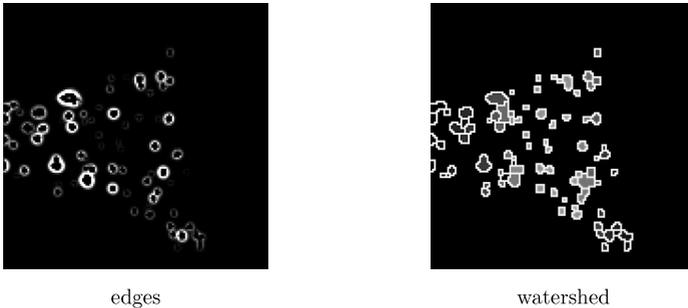


Figure 2.11: Edge highlighting and watershed segmentation. The edges are computed from the right image shown in Figure 2.10. Images are contrast enhanced.

2.2.3 EVALUATION

In the following sections the accuracy and precision of the presented algorithm are assessed. The tests are based on synthetic benchmark images of diffraction-limited objects for which the ground truth is known. Synthetic images were created using the imaging and object model specified in Equations 2.26 and 2.25, that is, the inferred model was valid. The benchmarks therefore quantify the sensitivity of the algorithm to experimentally controllable quantities, such as the level of noise in the images, but not the effect of departure from the model assumptions¹. Objects of different shape are used in order to test which shape features can be robustly recovered from blurred, noisy images. Furthermore, the sensitivity to the regularization parameters will be shown in an illustrative example. Before the actual benchmark results are discussed in Sections 2.2.3.2 and 2.2.3.3, the protocol for generating synthetic data will be explained in Section 2.2.3.1.

¹For the type of experimental data analyzed in Section 2.2.4, the validity of the model is questionable. Inspection of the residual of the fitted images, however, allows assessing to what extent different models could potentially improve the fit.

2.2.3.1 Generation of Synthetic Images

Synthetic fluorophore densities are generated from ground truth outlines according to Equation 2.25. The constant fluorophore density c is set to 200 and a background level of $b = 20$ is added (see Figure 2.12A). Imaging is simulated by a convolution of the objects with a measured point spread function of a typical experimental setup (full width at half maximum was 322 nm), downsampling to original pixel size, and addition of Poisson-distributed noise (Figure 2.12B and C). In order to add the proper physical noise level, the (gray-scale) image intensities $I_{i,j}$ are transformed to expected photon counts \hat{I} using the linear function $\hat{I}_{i,j} = \eta I_{i,j}$. Noisy images are then obtained by sampling a new value for each pixel (i, j) from a Poisson distribution with parameter $\hat{I}_{i,j}$ and transforming back to gray-scale intensities. By varying η , the signal-to-noise ratio $\text{SNR} = (c - b)/\sigma_c$ can be varied, where σ_c is the noise level in the center of the objects.

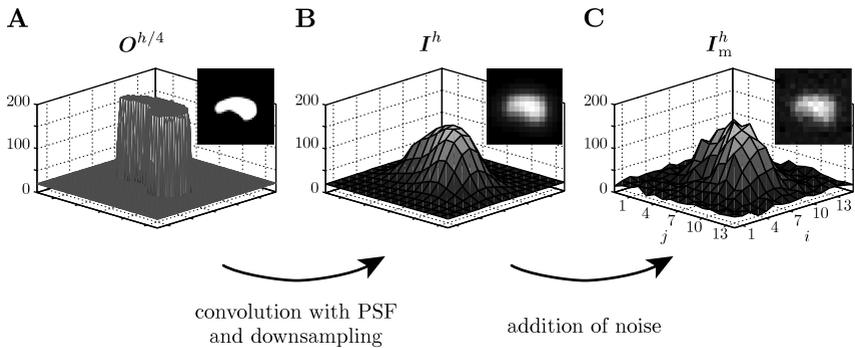


Figure 2.12: Generation of a synthetic benchmark image: A fluorophore density is generated from the ground truth outline (A). The image is generated by convolution and downsampling (B), and addition of noise (C).

2.2.3.2 Quantitative Analysis

The performance of the outline reconstruction algorithm is assessed on two sets of synthetic data. The first set contains images of a circular object with a diameter of 500 nm, the second set contains images of a 200 to 400 nm wide and 900 nm long pear-shaped object (see Figure 2.12). Images are generated as described in the previous section, with signal-to-noise ratios ranging from 7 to 56. $N = 250$ independent realizations of the noise process were generated for each true object and SNR.

The accuracy and precision of the fitted position (Figure 2.13), total intensity (Figure 2.14), and shape (Figure 2.15) are quantified. The position error is defined as the difference between the true and detected x -position of the intensity centroid. This definition is used in favor of the classical Euclidean distance in (x, y) because it enables correlating shape asymmetries with the position errors in the different directions. The total intensity error is given by the difference in total intensity (sum of all $O_{i,j}$ enclosed by the outline) between the reconstructed object and the true synthetic object, divided by the total intensity of the true synthetic object. The shape error is defined as the sum of non-overlapping areas of the true and fitted outlines, normalized by the area enclosed by the true outline. In Figures 2.13 to 2.15 the \pm standard error interval is shown in the bias plots (lines without markers) in order to enable visual assessment of the significance of the estimated bias. Standard errors are given by the estimated standard deviation divided by \sqrt{N} .

For both shapes, precision and accuracy of the measured position are in the range of a few nanometers (Figure 2.13). The errors in the y -direction are comparable for the pear-shaped object and identical for circular shapes (data not shown). The position bias for the circular shape (dashed lines) is always within the ± 1 standard error interval and hence not significant. A small systematic position bias can be observed for the pear-like shape (solid lines). This bias is caused by the internal energy of the outline and the imaging process, which limit the curvature of reconstructed outlines. For shapes with asymmetric curvature, this systematic under-estimation of curvature is translated to a small bias in the estimated position. In case of the pear-like object, this bias is towards the less curved side, because

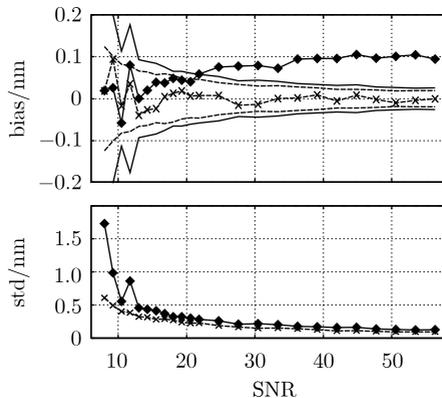


Figure 2.13: Position error. Accuracy (bias) and precision (standard deviation) of the reconstructed position of a pear-shaped (solid lines with diamonds) and a circular (dashed lines with crosses). Lines without markers delimit the ± 1 standard error interval.

the extent of the object on the more curved side is under-estimated.

For both shapes the standard deviation of the relative total intensity error drops below 5% for SNRs larger than 10 (Figure 2.14). The bias, however, is larger than the standard error, almost always negative, and converges to about -1% . A part of this bias can be attributed to the stretching term E_s in Equation 2.21, which favors shrinking of outlines and thereby decreases the integrated object intensity function. The estimation of the constant fluorophore density c is a further source of bias. This bias, however, increases the estimated total intensity, which partly counteracts the bias due to shrinking. While the bias caused by shrinking does not drop to zero (see Figure 2.15), the bias caused by the fluorophore density estimation does. This explains why the observed overall bias does not decay monotonically: At an SNR of ≈ 40 both effects are balanced, yielding an overall bias of zero. At larger SNRs the bias due to the fluorophore density drops to practically zero and the non-vanishing shrinking bias becomes again apparent.

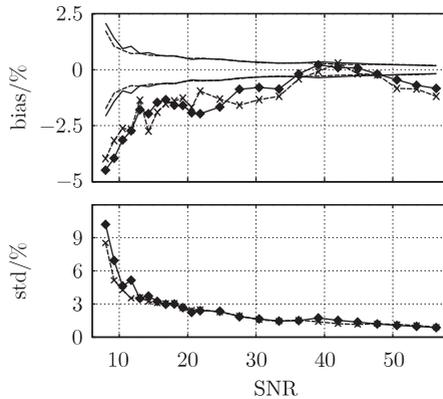


Figure 2.14: Intensity error. Accuracy (bias) and precision (standard deviation) of the reconstructed intensity of a pear-shaped (solid lines with diamonds) and a circular (dashed lines with crosses). Lines without markers delimit the ± 1 standard error interval.

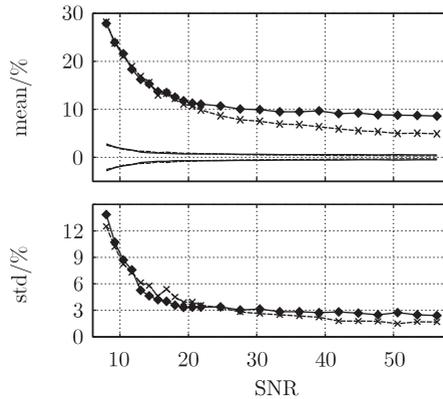


Figure 2.15: Shape error. Accuracy (bias) and precision (standard deviation) of the reconstructed shape of a pear-shaped (solid lines with diamonds) and a circular (dashed lines with crosses). Lines without markers delimit the ± 1 standard error interval.

The means and standard deviations of the shape errors of both shapes converge to values of less than 10% and 3%, respectively (Figure 2.15). The bending and stretching terms E_b and E_s of the internal energy prevent the mean shape error from converging to zero, and the systematic under-estimation of high curvature causes an additional bias for the more complex pear-shaped object. Nevertheless, we observe that the reconstructed outlines visually reproduce well the essential characteristics of the true outlines, even at SNRs below 15 and therefore shape errors of more than 10% (Figure 2.16).

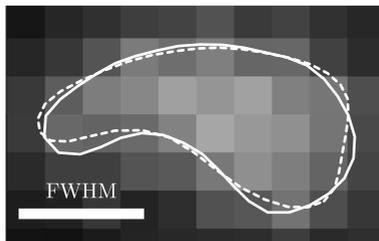


Figure 2.16: Reconstruction of an outline from a noisy image. Solid line: ground truth; dashed line: reconstruction; scale bar: full width at half maximum of the point spread function, in this case $\text{FWHM} = 322 \text{ nm}$. The SNR in this example is approximately 14.

2.2.3.3 Effect of Regularization Parameters

As discussed in the previous section, the regularization of the outline by the internal energy (E_s and E_b) introduces a bias toward less complex outlines in the reconstruction. This is because regularization acts to suppress insignificant undulations in the outlines. Such undulations may reduce the errors of the reconstruction, but are not robustly estimated and hence not trustworthy. This is similar to the problem of over-fitting as discussed in the context of classification in Chapter 1.

Adjusting the weight of the bending energy β (Equation 2.21) allows trading the accuracy of the outline estimation against its robustness by limiting

the aforementioned insignificant undulations of the contour. Low values lead to a higher noise sensitivity (less regularization), but allow better estimation of high curvatures, for example at sharp corners of an object. In order to qualitatively assess this trade-off, outlines are reconstructed from synthetic images of a triangle generated as described above (Figure 2.17). As expected, one can observe that low SNRs favor high values of β , and vice versa. The outline reconstruction is robust over two orders of magnitude of β . Only the most extreme case ($\beta = 0.02$, SNR = 5) exhibits significant shape instabilities. The stretching stiffness α has much less influence on the final contour (not shown). Higher values lead to faster convergence of the algorithm and better escape from local minima. At the same time, however, they bias the outlines to shorter, more contracted contours. For the data used in the present benchmarks $\alpha = 0.005$ is empirically found to be a good compromise between the ability to overcome local energy minima, speed of convergence, and accuracy.

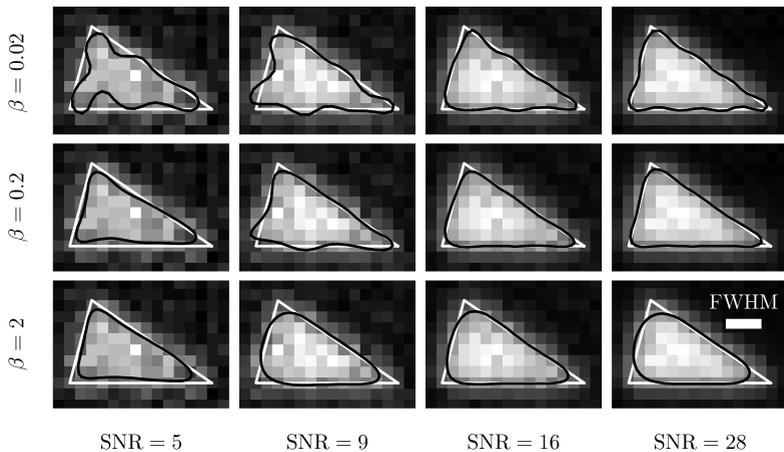


Figure 2.17: Regularized reconstruction of a highly curved outline as a function of the bending energy weight β and signal-to-noise ratio ($\alpha = 0.005$ in all cases). White triangles depict ground truth, black lines the reconstructed outlines. scale bar: full width at half maximum of the point spread function, in this case FWHM = 322 nm.

In this section, the utility of the present algorithm is demonstrated on fluorescence microscopy images of different intra-cellular structures, namely the Golgi complex and endosomes. The demonstrations can be divided into two groups: In the first group the intention is to illustrate the algorithm's deconvolution performance on concrete examples. Therefore, the segmentations, model images, and residuals are shown. This allows visually assessing the validity of the model assumptions. This case includes both Golgi (Section 2.2.4.1) and endosome (Section 2.2.4.2) images. The second group of demonstrations shows the value of the reconstruction algorithm for morphometric measurements of endosomes. Three separate cases are considered: estimation of shape feature distributions under different experimental conditions (Section 2.2.4.3), quantification of dynamic shape changes during endosome fusion (Section 2.2.4.4), and quantification of correlations between endosome outlines and virus particle locations during infection of cells with different virus strains (Section 2.2.4.5).

2.2.4.1 Implicit Deconvolution of Golgi Images

Figure 2.18 shows an image of the Golgi complex in a HeLa cell labeled by fluorescent giantin antibodies. The Golgi is a complex-shaped intra-cellular organelle composed of membrane stacks of about $5\ \mu\text{m}$ size. The cells considered contained a single Golgi, localized around the nucleus. Due to optical sectioning, however, it may appear as a few disconnected entities on the order of $5\ \mu\text{m}$ in size. The same image was also used to demonstrate active mask segmentation [136]. Very similar images were used to show a new mechanism for the regulation of Golgi size [55]. Such applications clearly benefit from bias reduction of size estimates by incorporating deconvolution into the segmentation procedure. I therefore show how a coarse, pixel-level segmentation can be refined by the present implicit deconvolution method.

The outline refinement algorithm is initialized by a rough manual segmentation obtained from Figure 11D in Reference [136]. Since no information about the point spread function was available, it is modeled as a Gaussian

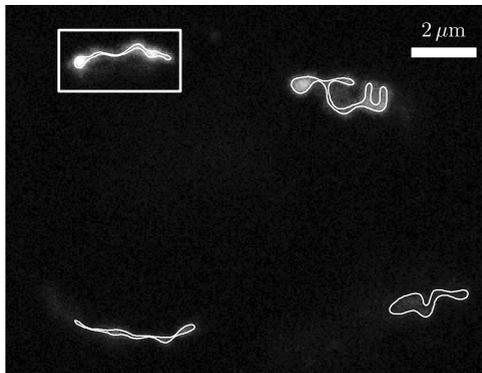


Figure 2.18: Reconstructed outlines of the Golgi of a single HeLa cell. The four disconnected outlines belong to the same Golgi. A magnification of the area in the white rectangle is shown in Figure 2.19.

with $\sigma = 150$ nm. This is a conservative choice for the imaging set-up used (spinning disk confocal, NA = 1.4, oil immersion).

The outline reconstruction algorithm is run until convergence as described. The final outlines (Figure 2.18 and magnification in Figure 2.19A) capture well the morphological characteristics of the Golgi. The estimated outline in Figure 2.19A shows no obvious signs of over-fitting. The corresponding model image \mathbf{I} is visually remarkably close to the measured image (Figure 2.19B). Nevertheless, the residual error $\mathbf{I}_m - \mathbf{I}$ (Figure 2.19C) shows that the model image tends to be too bright in the center of the object and under-represents the blur around it. This is probably caused by a violation of assumptions of the imaging model. For example, if the object is out of focus, the in-focus point spread function is not valid. In this case, a wider point spread function could potentially decrease the observed difference. Such adaptation of the point spread function would amount to performing *blind deconvolution*.

In some cases, the outlines fail to follow all features of the data. For example, the lower left outline in Figure 2.18 appears to be too short. This can be explained with a very low signal that is not sufficient to generate a strong image energy gradient. The internal forces are therefore dominat-

ing, which causes the outline to shrink to brighter regions.

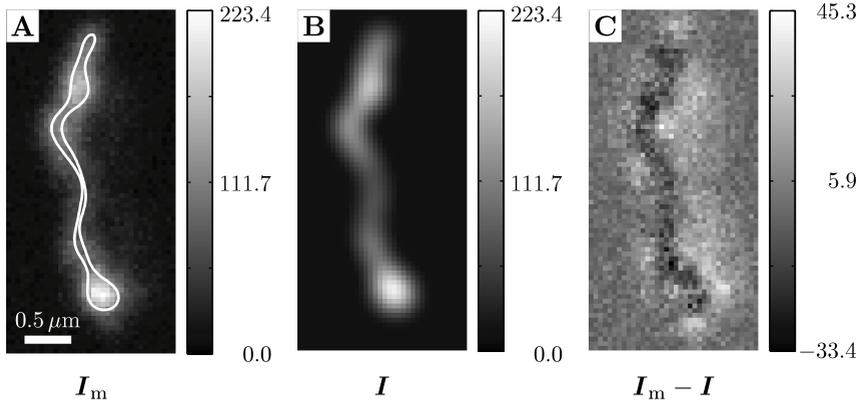


Figure 2.19: Magnification of the reconstructed outline of a part of the Golgi complex in a HeLa cell (A), model image (B), and residual (C). Note that a different gray value scale is used in C.

2.2.4.2 Implicit Deconvolution of Endosome Images

The second illustration of the algorithm’s performance considers live HER 911 cells expressing EGFP-tagged Rab5, a protein marker for endosomes. Endosomes are dynamic lipid-bounded organelles that are formed by invaginations of the plasma membrane [104]. With diameters of about 500 nm, endosomes are much smaller than the Golgi and they appear more compact. Endosomes are unevenly distributed in the cell and cover a wide range of sizes and intensities. In some cases, individual endosomes are very close to one another, which makes the grouping of close objects prior to simultaneous refinement very important.

Initialized by watershed segmentation (see Section 2.2.2.7), the present algorithm reconstructs the endosome outlines as shown in Figures 2.20 and 2.21A. The point spread function of the microscope (spinning disk confocal, NA = 1.35, oil immersion, full width at half maximum: 322 nm) was mea-

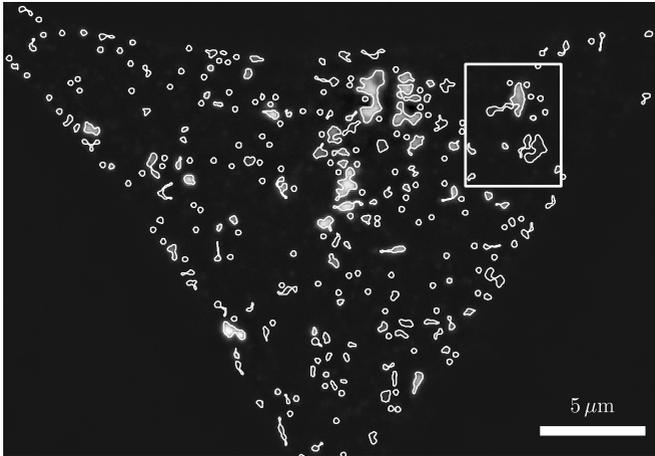


Figure 2.20: Reconstructed outlines of endosomes in a live HER 911 cell. A magnification of the area in the white rectangle is shown in Figure 2.21.

sured from images of sub-diffraction objects as described in Section 2.1.6.1. Figure 2.21A shows a magnification of complex-shaped endosome outlines in close vicinity. The outlines follow well the subjective contours in the images, even for very dim objects. The visual correspondence between the real and the model image (Figure 2.21B) is remarkable. Unlike in the Golgi case, there is no clear trend in the residual error and the error is smaller overall (Figure 2.21C), highlighting the advantage of using the true, measured point spread function. Except for slight over-estimation of the central intensity of the large object on the top, the residual error is dominated by detector noise. While this does not prove the validity of the model assumptions for endosome images, it shows the limited potential of more complex models to increase the quality of the fit and to be thus favored in statistical model selection.

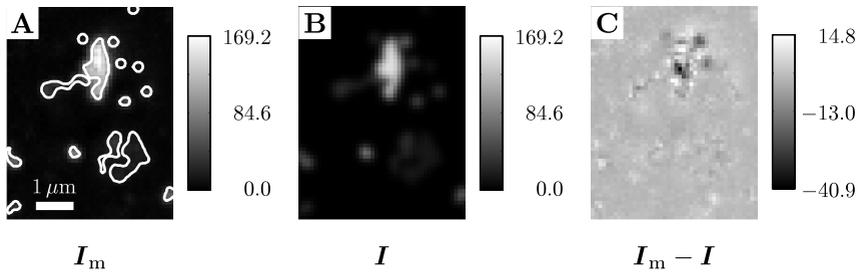


Figure 2.21: Magnification of reconstructed outlines of close-by endosomes in a live HER 911 cell (A), model image (B), and residual (C). Note that a different gray value scale is used in C.

2.2.4.3 Shape Feature Distributions of Endosomes

So far the applicability of the outline reconstruction algorithm has been demonstrated on a few sample images. In this and the following two sections I will show the benefits of refining endosome outlines in a larger explorative experimental study.

Endosomes are highly dynamic structures involved in processes such as endocytosis and intra-cellular trafficking. The complexity of endosomal shapes has previously been demonstrated using cryo-electron microscopy and tomographic reconstruction in fixed samples at high spatial resolution [60]. In live cells, however, the static and dynamic morphology of endosomes has not been characterized so far, partly due to difficulties pertaining to acquiring and analyzing light microscopy images of such small objects. Endosome morphology could actively or passively be related to endocytic functions. Therefore, accurate quantification of endosome shape features is of potential interest.

A large set of live HER 911 cells stably expressing the small GTPase Rab5 tagged with enhanced green fluorescent protein (911-EGFP-Rab5) [121] are analyzed. Virus particles are used as model cargo to show a potential interaction between endocytic cargo and endosome morphology. In total 416 cells are analyzed: 201 cells are infected with Adenovirus serotype 2

(Ad2), and 184 cells are infected with its temperature-sensitive mutant TS1. In addition, control images of 31 non-infected cells are recorded and analyzed.

Based on the reconstructed endosome outlines, a set of descriptive shape features is computed and correlated with the presence of virus particles in whole cells and individual endosomes. From an originally larger set of features, four have proven most informative: the *area*, *total intensity*, *eccentricity*, and *concavity* of individual endosomes.

- **Area** is computed as the area enclosed by the outline Θ .
- **Total intensity** is area multiplied by the estimated fluorophore density c .
- **Eccentricity** is computed from the radius of gyration tensor G of the surface enclosed by Θ . It is defined as the ratio between the largest and the smallest Eigenvalue of G .
- **Concavity** is computed from the area A_{ch} of the convex hull of an outline. It is defined as the ratio $(A_{\text{ch}} - A)/A_{\text{ch}}$, where A is the area enclosed by the outline.

Since the present method is completely automatic, it enables processing large amounts of data in an unbiased and reproducible way. This leads to increased statistical significance of the final results. Furthermore, the algorithm implicitly corrects for the microscope optics, that is, it implicitly deconvolves images. Therefore, computation of shape the features is robust and the bias is reduced. The measured distributions of the four shape features are shown in Figure 2.22 for the control and TS1-infected cells. Shape feature distributions for Ad2-infected cells are similar to those of TS1-infected cells and are not included in the plots for clarity.

Using these distributions, the question whether viruses preferentially occupy endosomes with specific shape features is addressed. Each endosome is assigned to one of two groups, depending on whether it enclosed at least one virus particle or not. Comparison of the shape feature distributions of the virus-containing set (“pos.”, green dash-dotted lines) with the distributions of all Rab5 endosomes (“all”, blue dashed lines) shows

that viruses are more frequently found in large endosomes with high fluorescence intensity. This preference is significantly stronger than what one would expect under random placement of the viruses in the cell (“rand. pos.”, dotted red lines), ruling out a linear dependence of the frequency of occupation on endosome area. Interestingly, virus-containing endosomes have a below-average eccentricity, but slightly increased concavity² (Figure 2.22C and D) compared to the population of all endosomes in the cell. Using the randomized virus positions, a clear trend toward higher concavities is observed, which is consistent with the fact that the complex shapes of outlines can be better resolved on larger scales.

In order to show the benefit of the present outline reconstruction algorithm, the obtained shape feature distributions are compared to those from the outlines of the unrefined watershed segmentation (see Section 2.2.2.7). For the latter, good correspondence between the segmented outlines and reality is not guaranteed, since no imaging model is used for their construction.

For area and total intensity (data for unrefined outlines not shown), the shape feature distributions show the same trend when comparing TS1-containing endosomes to the population of all endosomes, although the trend is less significant. For concavity, the results look very different (Figure 2.23). In unrefined outlines (right peak), high concavity (>0.3) is virtually non-existent. Instead, there is a pronounced peak around 0.1, which can be attributed to the step-like, non-smooth shapes of the outlines. This significantly reduces the information content of the observations. The inset images in Figure 2.23 show a fusion of two endosomes (as I will discuss in Section 2.2.4.4), with overlaid refined (outer images) and unrefined (inner images) outlines. During the fusion, concavity temporarily increases from 0 to 0.2. On unrefined outlines, an uninformative change from 0.09 to 0.12 is measured (horizontal distance arrows in Figure 2.23). Detection of shape change events, such as fusion, may thus depend on the use of properly refined outlines.

The benefits of using the present outline reconstruction algorithm can be formally shown by applying statistical tests to the shape feature distributions. A two-sided Kolmogorov-Smirnov test is performed on the null

²Although visually very small, the difference between the distributions is statistically significant owing to the large number of samples.

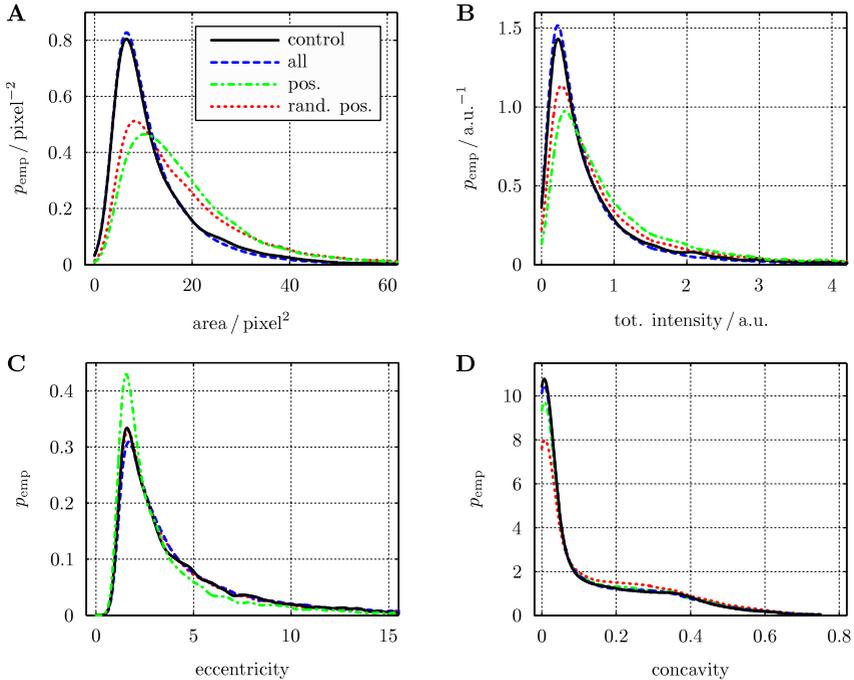


Figure 2.22: Shape features of Rab5-positive endosomes in HER-911 cells. Empirical distributions of endosome area (A), total intensity (B), eccentricity (C), and concavity (D) are shown for reconstructed endosome outlines. Each panel shows the distributions for: all endosomes in non-infected control cells (solid black lines, $N = 4581$), all endosomes in cells infected with TS1 (dashed blue lines, $N = 31351$), endosomes containing at least one TS1 particle (dash-dotted green lines, $N = 4119$), and endosomes containing at least one TS1 particle with randomized position (dotted red lines, average $N = 3458$). Distributions were estimated using a Gaussian kernel density estimator.

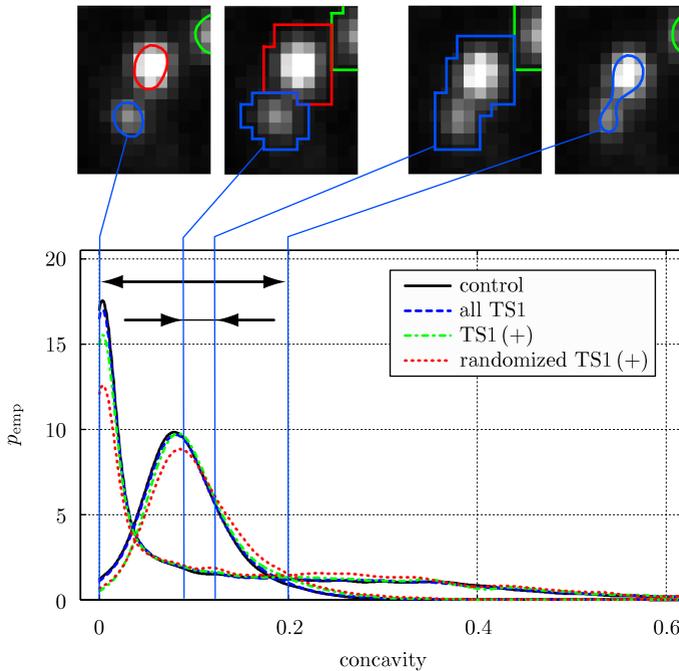


Figure 2.23: Comparison of concavity distributions from unrefined (right peak) and refined (left peak) outlines. Inset images compare the shapes immediately before and after a fusion event of two endosomes (will be discussed in Section 2.2.4.4). Using refined outlines, a more pronounced change in concavity is observed (horizontal distance arrows).

hypothesis that two shape feature distributions are identical. All possible combinations of the three cases of TS1-infected cells, Ad2-infected cells, and control virus are tested. *All* endosomes in each cell are considered, not only those that actually contain virus(es). This amounts to investigating whether the presence of virus particles in cells has an overall effect on endosome shape. These tests are not specific to a certain mechanism that could cause shape differences. Therefore, they at best allow concluding that a signal is present that is not accounted for by the null hypothesis.

The results of the pair-wise comparisons are summarized in Figure 2.24. The p-values are color-coded according to the significance level and the signs indicate the directions of the shifts of the means. Since large amounts of data were used, the null hypothesis could be rejected in some cases, despite the small visual differences between the distributions (Figure 2.22, “all” versus “control”). By comparing test results between features and combinations of cells, the self consistency of the tests can be assessed. Test results for refined outlines (Figure 2.24A) show a consistent picture: TS1 infection significantly changes area, total intensity, and eccentricity (red and orange colors in first row), but Ad2 infection does not (green and yellow colors in second row). The third row, comparing TS1 against Ad2, confirms this result. When using unrefined outlines, the test outcomes are not self consistent (Figure 2.24B) and do not convincingly support any conclusion. For example, area and total intensity are correlated quantities and the color patterns of these two columns should be similar. While this is the case when using refined outlines, the patterns for unrefined outlines look opposite.

2.2.4.4 Fusion of Endosomes

Using time-lapse microscopy of live cells, dynamic morphological changes of endosomes can be observed. The required high temporal resolution can only be achieved at the expense of a low signal-to-noise ratio, which challenges outline reconstruction. In order to capture the dynamics of endosome fusion events, $2.2\ \mu\text{m}$ thick *z*-stacks, comprising 4 images each, were recorded in the cell periphery at high frequency (25 stacks/s) on a spinning disc confocal microscope. After acquisition, the image stacks

A	area	tot. int.	ecc.	conc.
TS1 / Ctrl	$8.6 \cdot 10^{-4} (-)$	$2.0 \cdot 10^{-6} (-)$	$2.5 \cdot 10^{-4} (+)$	$5.1 \cdot 10^{-1} (+)$
Ad2 / Ctrl	$8.4 \cdot 10^{-1} (+)$	$1.4 \cdot 10^{-2} (-)$	$3.6 \cdot 10^{-2} (+)$	$8.4 \cdot 10^{-1} (+)$
TS1 / Ad2	$1.9 \cdot 10^{-10} (-)$	$9.4 \cdot 10^{-9} (-)$	$3.0 \cdot 10^{-14} (+)$	$4.6 \cdot 10^{-1} (+)$

B	area	tot. int.	ecc.	conc.
TS1 / Ctrl	$2.4 \cdot 10^{-2} (-)$	$6.7 \cdot 10^{-6} (-)$	$5.4 \cdot 10^{-5} (+)$	$1.5 \cdot 10^{-1} (+)$
Ad2 / Ctrl	$3.9 \cdot 10^{-2} (+)$	$1.2 \cdot 10^{-3} (-)$	$1.2 \cdot 10^{-1} (+)$	$7.4 \cdot 10^{-1} (+)$
TS1 / Ad2	$4.2 \cdot 10^{-6} (-)$	$2.5 \cdot 10^{-2} (-)$	$9.0 \cdot 10^{-10} (+)$	$1.1 \cdot 10^{-2} (+)$

p-value	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^0
---------	-----------	-----------	-----------	-----------	-----------	--------

Figure 2.24: Self consistency of statistical tests on the shape feature distributions. The p-values of two-sided Kolmogorov-Smirnov tests of the null hypothesis of identical distributions are shown (color codes significance level). First row: all endosomes in TS1-infected cells (“all” in Figure 2.22, $N = 31351$) versus control cells (“control” in Figure 2.22, $N = 4581$ endosomes); second row: all endosomes in Ad2-infected cells ($N = 34323$) versus control; third row: TS1-infected cells versus Ad2-infected cells. Signs mark the direction of the trend with a (+) denoting that the first condition has a larger mean feature value than the second one, and a (-) the opposite. A: refined outlines; B: unrefined outlines.

were reduced to single images by maximum projection [120]. Selected frames of a movie are shown in Figure 2.25A. Outlines of the endosomes were reconstructed in each frame of the movie and individual detections were linked over time [125], yielding endosome trajectories.

Dynamics of endosome shapes and fusion events are quantified by the time evolution of the area, total fluorescence intensity, eccentricity, and concavity of the involved endosomes (see Section 2.2.4.3 for definitions of the shape features). Because the algorithm implicitly corrects for the microscope optics, computation of shape the features is robust and unbiased (see Section 2.2.4.3).

The dynamics of the endosome shape features computed from refined outlines are shown in Figure 2.25B. Pronounced changes in endosome shape features are associated with topological and morphological changes of the endosomes. At 0.92 s, a large endosome (Figure 2.25A, red outline) rapidly approached an immobile one (blue outline) until the two structures could no longer be resolved. At 1.08 s, the algorithm detected only a single outline. In the following, the endosome remained stationary before it coherently displaced along a linear track as a single entity (4.08 s). We conclude that the merged object represents a fused endosome with different dynamic behavior. A transient peak in concavity and eccentricity during about 120 ms marked the fusion event (Figure 2.25B, blue lines). As expected, the area increased upon fusion, and the total fluorescence intensity of the fused endosome matched the sum of the two pre-fusion intensities (Figure 2.25B, red line and blue line). In contrast, the features of a stationary endosome remained constant (green lines).

Changes in area and intensity, together with the sharp peaks in concavity and eccentricity, can be used as hallmarks of fusion events. Note that fusions are very fast events and that, therefore, high temporal resolution is required to detect the accompanying transient changes in morphological features. The outline reconstruction algorithm enables automatic, unbiased, and reproducible estimation of shape features at high resolution. It therefore provides a prerequisite for detecting and analyzing fusion events.

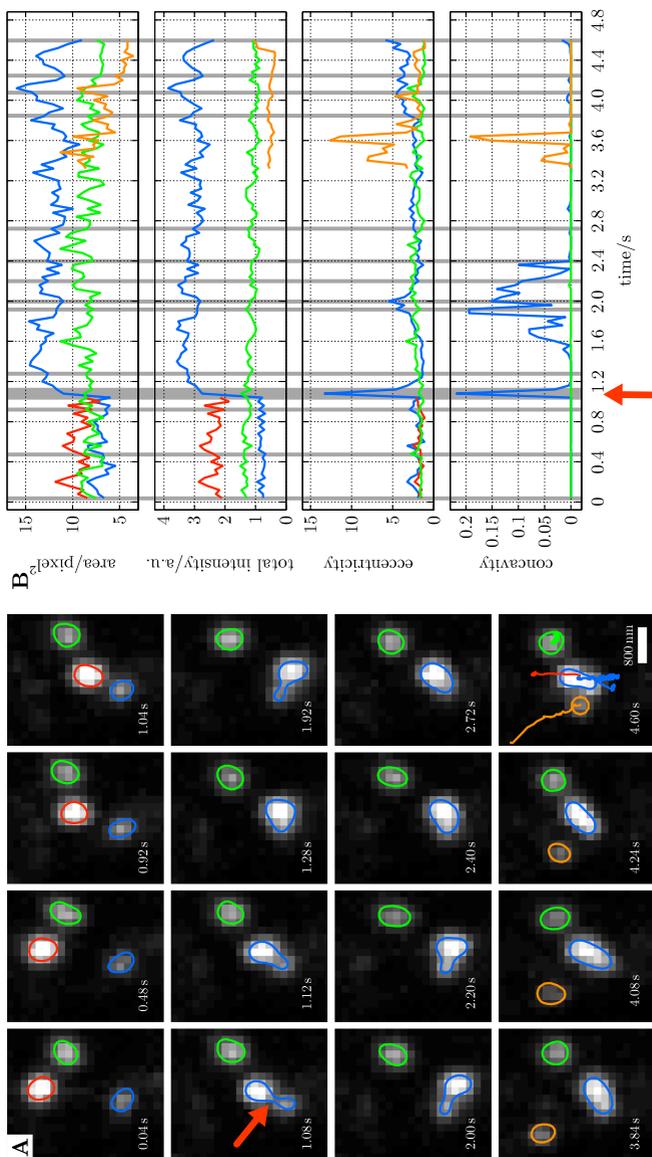


Figure 2.25: Dynamic shape changes of fusing Rab5-positive endosomes. (A) A region near the periphery of an HER 911 cell, imaged at 25 Hz. Selected frames with overlaid outlines are shown. Trajectories of the endosomes are shown in the last frame. (B) Area, total intensity, eccentricity, and concavity are plotted for the indicated objects in (A) during a period of 4.8 seconds. Gray vertical lines mark the time points of the frame images in (A).

2.2.4.5 Co-localization of Viruses and Endosomes

The last application considers trafficking of virus particles through the endosomal network. Virus trafficking can be monitored by quantifying the *co-localization* of endosomal and viral markers [13]. The same data as in Section 2.2.4.3 is considered: HER 911 cells infected with either Adenovirus serotype 2 or its temperature-sensitive mutant TS1. TS1 visits early endosomes like wild-type Ad2, but fails to escape to the cytosol and is instead delivered to late endosomes and lysosomes [43]. Figure 2.26 shows the two color channels overlaid for a cell imaged 10 min post infection. The overlay shows that some viruses seem to co-localize with Rab5. The degree of co-localization and its significance, however, can not be estimated by eye.

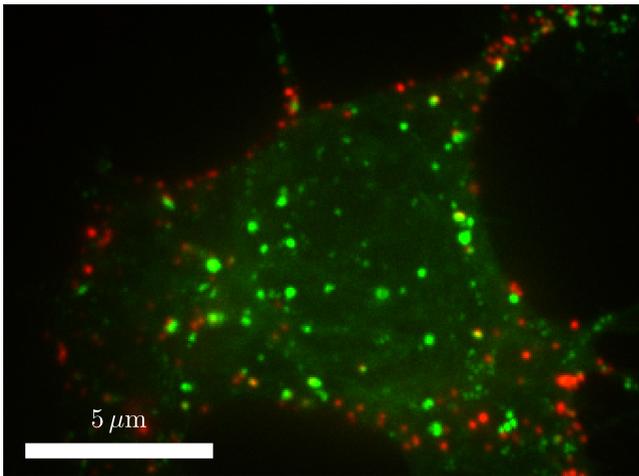


Figure 2.26: Two-color fluorescence microscopy image of endosomes (green: EGFP-Rab5) and virus particles (red: TS1-atto647).

By reconstructing endosome outlines and estimating of virus positions, the image data are transformed to sets of discrete objects. In this representation, “co-localization” is defined as the fraction of viruses that are enclosed by an endosome outline. As in Section 2.2.4.3, the results are compared

to those obtained using unrefined outlines as found by watershed segmentation. While the unrefined outlines are certainly correlated with the true outlines, they are likely to be biased and more sensitive to variations in individual pixel intensities. Figure 2.27 supports this argument: Considering the blurring introduced by the microscope point spread function, the unrefined outlines are clearly too wide.

Figures 2.28A and B show histograms of distances of viruses to the nearest endosome outline in the cell shown in Figure 2.27. The distribution from unrefined outlines (A) has a broader peak and is shifted toward smaller distances when compared to refined outlines (B). In order to test how these differences affect the final results, the co-localization scores of virus particles and Rab5-positive endosomes are determined using either refined or unrefined outlines. It is known that Ad2 escapes from an endosome soon after internalization [53]. We thus expect little co-localization of Ad2 with Rab5. In contrast, the signaling-incompetent mutant TS1 is known to reside in early endosomes and lysosomes during the first hour of entry into cells [28, 52, 43], probably yielding a transient, but long-lasting, increase in co-localization with Rab5.

Colocalization scores were determined for 201 (Ad2) and 184 (TS1) cells, each imaged at a single time point between 0 and 45 min post infection. In order to estimate the degree of unspecific co-localization, controls with randomized virus positions were performed. The means and standard deviations of the co-localization scores are estimated using a Nadaraya-Watson kernel estimator with bandwidth $h = 3$ min [132] and plotted versus time. For the kernel estimation, individual co-localization scores were weighted proportionally to the number of viruses detected in the respective cell. The time courses of Ad2 and TS1 co-localization with Rab5-positive endosomes are shown in Figures 2.28C and D. The level of unspecific co-localization that would be expected under random placement of viruses was approximately 1 to 5% for refined outlines and 2 to 10% for unrefined outlines.

The mean co-localization score for Ad2 remains fairly constant on a comparably low (but above background) level of less than 10% when using refined outlines (Figure 2.28D, lower line). The score for TS1 (Figure 2.28D, upper line) shows an increase from 12 to 20% between 10 and 20 min post infection and then remains high throughout the observation period. For both

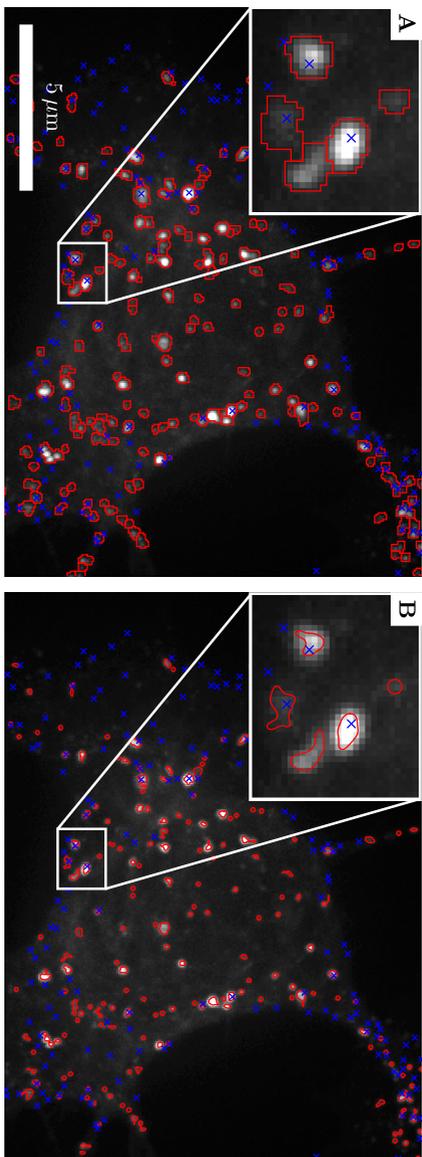


Figure 2.27: HER-911 cell expressing EGFP-Rab5 imaged 10 min post infection with overlaid endosome outlines (red lines) and positions of TS1 virus particles (blue crosses); (A) unrefined, (B) refined outlines.

viruses, the ± 1 standard deviation intervals (light gray areas) are narrow and do mostly not overlap (dark gray areas). In summary, Ad2 and TS1 show co-localization signatures with refined outlines that are significantly distinct in magnitude and dynamics and that are in good agreement with prior observations [53, 52, 43].

This picture is much less clear when using unrefined outlines. While the increase around 15 min post infection is still apparent for TS1 (Figure 2.28C, upper line), the score for Ad2 (Figure 2.28C, lower line) increases faster from 0 to 45 min post infection than when using refined outlines (Figure 2.28D). Compared to the results based on refined outlines, larger means and standard deviations are observed (gray shaded areas) for both viruses. The ± 1 standard deviation intervals significantly overlap during the entire observation period (dark gray areas), which makes it more difficult to distinguish between the two viruses with fewer data points. In summary, the qualitative and quantitative trends seen in Figure 2.28C are less clear than the results in Figure 2.28D, which, owing to their agreement with prior observations, must be considered more trustworthy. This demonstrates the benefits of using the present outline refinement algorithm in biological studies.

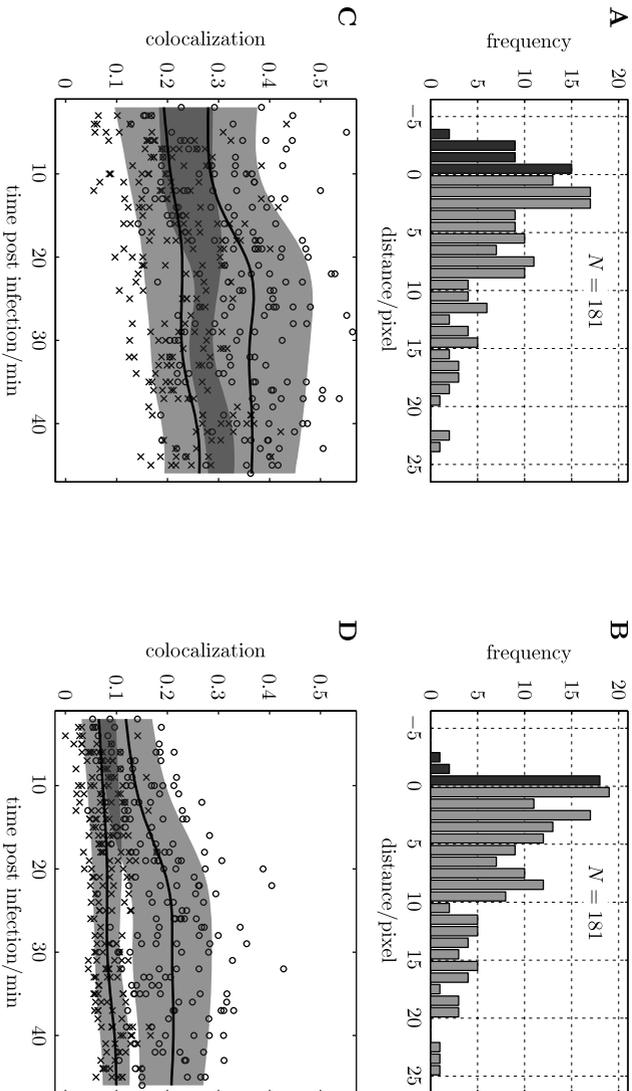


Figure 2.28: Virus–endosome co-localization. (A and B) Histograms of distances of viruses to the nearest point on endosome outlines for the cell in Figure 2.27. Negative distances correspond to viruses inside endosomes. (C and D) Mean (black lines) and $\pm 1\sigma$ interval (shaded areas) of the co-localization scores versus time post infection. Ad2 (lower line) and TSI (upper line). Scores for individual cells are shown as crosses (Ad2, $N = 201$, lower lines) and circles (TSI, $N = 184$, upper lines); (A and C) unrefined outlines, (B and D) refined outlines.

CHAPTER
THREE

QUANTIFYING INTERACTIONS BETWEEN
INTRA-CELLULAR OBJECTS

This chapter describes a collection of methods for statistical inference of interactions between intra-cellular objects from correlations in their spatial locations.

Quite generally one could state that cellular function results from the combined interactions of sub-cellular structures in space and time. The study of the spatial localization of sub-cellular structures has therefore attracted great attention [8]. Interactions typically manifest themselves through statistical dependencies in the spatial distributions of the involved structures. In this view one can define *interaction* as the collection of all effects that cause significant (above the level predicted by a null hypothesis) correlations in the positions of the participating objects. “Interaction” does then no longer refer to some specific process (for example protein–protein interaction), but is rather an abstraction that can be used to describe so far unknown processes.

Over the last decades, advances in fluorescent markers have enabled probing interactions of sub-cellular structures in the microscope, either directly or indirectly. The *direct approach* relies on experiments that generate a signal upon the proximity required for molecular interaction, for example Förster resonance energy transfer (FRET) or bioluminescence resonance energy transfer (BRET) [48, 68, 116]. The idea of these methods is that the energy of the excited state (see Section 2.1.1) of a donor chromophore, M_d , is non-radiatively transferred to an acceptor, M_a . M_d and M_a must have significantly different emission and excitation spectra, but the emission spectrum of M_d and the excitation spectrum of M_a need to overlap in order to allow resonance transfer. All wavelengths other than the emission spectrum of M_a are shielded from the detector. Detection of a photon then means that resonance energy transfer has occurred. Since resonance energy transfer requires a proximity on the order of nanometers, a molecular interaction is often concluded from a positive signal.

The *indirect approach* is based on *independently* imaging two populations of interest, and searching for clues of interaction in their spatial distributions. The paradigm that spatial proximity (or *co-localization*) is a hallmark of many types of physical and chemical interactions between sub-cellular structures is central to the indirect approach. If two or more structures interact, their spatial distributions should appear correlated. The reverse, however, is not necessarily true. Presence or absence of significant co-localization does not imply presence or absence of interaction. The reason is that co-localization depends on the specific interaction mechanism: An unobserved third structure may act as a confounding factor (in the statistical sense), making the observed structures appear co-localized even though they do not interact directly. Furthermore, one can imagine interaction mechanisms that lead to spatial distributions with correlations that are not captured by simple co-localization measures. Hence, the interaction has to be statistically *inferred* from the data.

Such inference, however, entails a trade-off between the objectives of pattern discovery and statistical detection power. According to these objectives, two complementary approaches to co-localization analysis can be distinguished: Intensity correlation methods capitalize on pattern discovery [27], whereas object-based methods [13] emphasize detection power.

Intensity correlation methods (Section 3.1) quantify correlations in the intensities of different color channels in individual pixels. Intensity correlation methods are straightforward to implement and use. The results, however, may be difficult to interpret since interactions need to be inferred from correlations in *intensity space*, which is sensitive to the blurring and noise inherent to microscopic imaging systems [2]. The intensity-based approach is limited to interactions on a spatial scale on the order of the resolution of the microscope.

Object-based methods (Section 3.2) quantify the spatial relationships between sets of discrete objects. This requires reducing the image to a set of geometric objects using, e.g., image segmentation or fitting of structure models. Object-based approaches infer interactions from correlations in *physical space*, which allows constructing intuitive and simple co-localization measures, such as counting the number of overlapping objects [13]. Nevertheless, they crucially depend on reliable methods to extract object descriptors from images. To what extent positive co-localization implies the presence of interactions, however, remains unclear.

In Section 3.2.2 a statistical model for the distances between potentially interacting objects is presented. Based on this model, the link between co-localization measures and interaction is established. The model allows addressing many frequently encountered data analysis tasks using standard statistical procedures, as demonstrated here on a large variety of examples that consider the virus–endosome dataset used in Chapter 2 (Section 3.2.5).

3.1 INTENSITY CORRELATIONS

The fundamental premise of intensity correlation methods is that spatial proximity of the imaged objects is manifested in correlations between the intensities of the different color channels in individual pixels of an image. In typical microscopy setups, the images are sampled with inter-pixel spacings of roughly 1/3 the width of the point spread function. Objects that are closer than the width of the point spread function will thus create correlated pixel intensities. In a scatter plot of pixel intensities in the two color channels, pixels close to co-localized objects will therefore appear

along a line close to the principle diagonal. The closer the objects and the larger their number, the higher the number of pixels with correlated intensities. Other pixels will, however, create a background signal of uncorrelated intensity values. These pixels may, depending on their relative frequency, obscure true correlations. Figure 3.1 shows an example of a scatter plot of pixel intensities from the virus–endosome image shown in Figure 2.26.

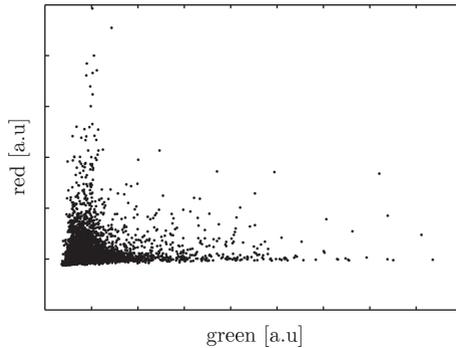


Figure 3.1: Scatter plot of single-pixel intensities of the image shown in Figure 2.26. x -axis: intensities in the image of Ad2 virus particles; y -axis: intensities in the image of Rab5-positive endosomes.

Since microscopy images result from an imperfect process, individual light sources appear blurred, noisy, and will have possibly different intensities depending on their location in the field of view (see Section 2.1.6). These effects directly impact the scatter plot of pixel intensities, which may hamper analysis [155, 13, 27]. Assume a red-green dual color image is given, where each color corresponds to a different type of object. The simplest measure of co-localization is the *sample Pearson correlation coefficient*:

$$r = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{R_i - \langle R \rangle}{s_R} \right) \left(\frac{G_i - \langle G \rangle}{s_G} \right), \quad (3.1)$$

where R_i and G_i are the red and green intensities of the i -th pixel, $\langle R \rangle$ and $\langle G \rangle$ the mean red and green intensity, and s_R and s_G the standard

deviation of the red and green intensities. Many improvements over this basic measure have been proposed that allow detecting a larger variety of correlations, and that are less sensitive to imperfections in the input images [13, 27].

According to the definitions in Reference [13], Figure 3.1 does not support the conclusion of any significant co-localization of endosomes with Ad2 virus particles. The Pearson correlation coefficient, for example, is negative. This may change if the images are carefully preprocessed in order to reduce the number of background pixels in the scatter plot. Nevertheless, the results remain difficult to interpret since the co-localization, and hence the potential interaction, of the imaged objects needs to be inferred from correlations in *intensity space*.

A peculiar property of intensity-based co-localization analysis is related to the optical resolution of the imaging device. The blurring and noise inherent to microscopic imaging systems influence the intensity correlation in a non-trivial way [96, 2]: Depending on the distance between the objects and on the noise level, an increase in optical resolution may both increase *or* decrease the observed intensity correlation. In that sense, the imaging device, rather than a design decision in the analysis, implicitly defines the length-scale on which interactions can be detected. A further complication is related to chromatic aberration [41] (see Section 2.1.6.2): Shifts between the images in the different color channels directly influence intensity correlations. Image registration has to be performed to correct for this artifact. Owing to their simplicity, intensity correlation methods are nevertheless popular tools for pattern discovery.

3.2 OBJECT-BASED INTERACTION ANALYSIS

Object-based approaches to inferring interactions work on the level of the locations of objects. The statistics used to describe interactions are therefore constructed from correlations in the same space (the physical space) in which the interactions take place. While this makes the statistics more intuitive, it requires a transformation from pixel images to discrete object representations.

The object-based approach is not necessarily limited to any particular length scale; a spatial scale is nevertheless assumed in practice. This is because many object-based co-localization methods rely on a hard threshold for the distances between objects in order to distinguish “co-localized” from “not co-localized” for each individual pair of objects [13]. The choice of distance threshold greatly influences the types of interactions that can be reliably detected. The actual physical or chemical interaction between sub-cellular objects may be of short temporal duration and the objects may quickly separate thereafter. In such situations, high thresholds can increase the detection power, but only at the expense of increased false-positive rates. When interactions take place over long distances, the choice of threshold implicitly determines a range limit of the analysis.

Apart from fixing the interaction scale *a priori*, using a hard distance threshold also implies a binary distinction of pair-wise distances: either they are below the threshold and hence the objects are assumed to interact – or they are not. A co-localization percentage thus corresponds to an indirect measure for the preference of “interaction” over “non-interaction”. This preference reflects the strength of the interaction. However, it also depends on the frequency of possible distances that the population of objects can assume. More specifically, the cellular context in which the interactions take place is a confounding factor. A high co-localization percentage can, for example, be observed in a cell with densely packed sub-cellular structures of interest, irrespective of their actual interaction strength. This artifact needs to be considered in statistical tests [154] or corrected for in order to construct an interaction score [83].

3.2.1 CLASSICAL CO-LOCALIZATION MEASURES

Object-based co-localization measures are typically constructed for two sets of objects $X = \{\mathbf{x}_i\}_{i=1}^N$ and $Y = \{\mathbf{y}_j\}_{j=1}^M$. These objects are located in a bounded region $\Omega \subset \mathbb{R}^n$ with boundary $\partial\Omega$ and dimensionality n (usually 2 or 3; see Figure 3.2). Each object i (j) is represented by a feature vector \mathbf{x}_i (\mathbf{y}_j) that comprises information about the object’s position and, if available, its size and shape. These feature vectors are extracted from image data by using image segmentation or fitting of structure models.

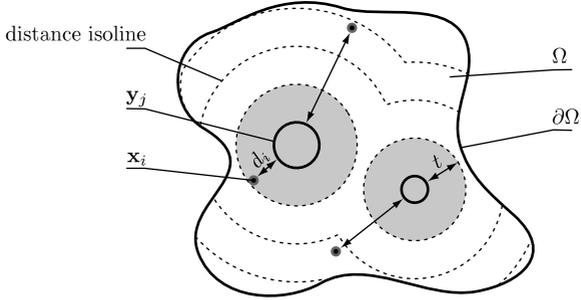


Figure 3.2: Illustration of co-localization analysis based on nearest neighbor distances (arrows) between point-like objects $X = \{\mathbf{x}_i\}_{i=1}^N$ (dots) and circular objects $Y = \{\mathbf{y}_j\}_{j=1}^M$ (solid circles). The expected co-localization in the absence of interactions, C_0^t , is proportional to the area enclosed by the t -isoline (gray region).

Suppose one wishes to investigate the interaction between the objects in X and Y , one can define for each \mathbf{x}_i the distance to the nearest neighbor (NN) in Y ,

$$d_i = \min_j \{d(\mathbf{x}_i, \mathbf{y}_j)\} . \quad (3.2)$$

The function $d(\cdot)$ is a suitable distance function in feature space, for example the Euclidean distance between point-like objects or the minimum distance between outlines of more complex objects. A nearest-neighbor distance distribution $p(d)$ can then be estimated from the set of distances $D = \{d_i\}_{i=1}^N$. $p(d)$ is the probability density function for observing a certain nearest-neighbor distance in Δd about d in the given cellular context *as caused by the interaction process*. The classical *overlap* or *nearest-neighbor-distance co-localization measure* C^t follows by counting [83]:

$$C^t = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(d_i < t) \xrightarrow{N \rightarrow \infty} \int_{-\infty}^t p(d) dd, \quad (3.3)$$

where $\mathbf{1}(\cdot)$ is the indicator function and t an application-specific distance threshold. The form of Equation 3.3 implies assumptions about how the objects in X and Y interact. The interaction process is considered to be

translation and rotation invariant since only the *distance* between interacting objects is taken into account, whereas the *location* and the *orientation* of the objects play no role. Based on this distance, only two categories of positions of the objects in X are distinguished: either they are sufficiently close to any object in Y to be considered interacting, or they are not. Furthermore, objects in X interact with at most one object in Y and they do not experience the presence of any \mathbf{y}_j unless they cross the distance threshold t . The choice of t reflects an (implicit) assumption about the length scale of the interaction to be detected.

When the distance threshold can not be justified by prior knowledge, the usual approach is to choose it in a systematic way. In the current formulation, the overlap measure is a descriptive statistic. This statistic, however, is not based on a statistical model, but rather on an ad-hoc definition. Therefore, it is not possible to directly apply standard procedures of model selection. Such procedures allow selecting the optimal model (in a statistical sense) out of a collection of models.

3.2.1.1 The Cellular Context

Inferring interactions from an observed co-localization measure C^t is not trivial since $C^t > 0$ does not necessarily imply any interaction between the objects. This is because spatial correlations can also be caused by confounding factors, such as the cellular context $\{\Omega, Y\}$. Even if the objects in X and Y do not interact, there is a non-zero probability that any possible distance in an interval Δd about d_i is observed. Y is arbitrarily chosen as a reference in order to compute the relative frequencies of possible distances (the *state density*) as:

$$q(d) = \lim_{\Delta d \rightarrow 0} \frac{\text{Prob}(d_i \in [d, d + \Delta d] | \text{“no interaction”}, Y)}{\Delta d}. \quad (3.4)$$

This density $q(d)$ is determined by the positions, sizes, and number density of the objects in Y (see Figures 3.2 and 3.3). Independent, uniformly random positions result in a relatively wide density $q(d)$ (Figure 3.3B). With regularly placed objects Y , large distances do not occur (Figure 3.3A). Clustering increases the frequency of long distances at the expense of short

distances (Figure 3.3C). Objects with large surfaces or a large number density give rise to shorter distances. In case there are interactions between the objects in X and Y , some of the possible distances are additionally favored over others, deforming the density $q(d)$ to $p(d)$. This deformation is typically more pronounced for stronger interactions. However, the interaction may also fail to yield a significant effect for certain functions $q(d)$: If, for example, the interaction takes place only at extremely short distances, the state space corresponding to these distances might be so small that hardly any effect is observed.

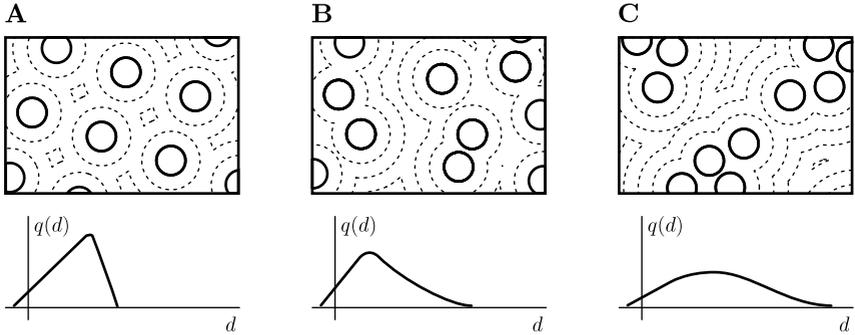


Figure 3.3: The cellular context determines the state density $q(d)$. For all distances d , $q(d)$ is proportional to the total length of the d -isoline (dashed lines) in Ω . (A)-(C) Effect of the positioning of the objects Y on $q(d)$, illustrating the influence of the cellular context..

The co-localization measure C^t is, therefore, not sufficient to separate the contributions from the cellular context and the interactions. Information about the interactions is only contained in the *deviation* from an expected base-level in the absence of interactions. This base level, say C_0^t , is the co-localization measure that would be observed under the hypothesis H_0 : “no interaction”. It can be obtained by letting $p(d) = q(d)$ and numerically evaluating the integral in Equation 3.3. In Figure 3.2 the gray regions represent the part of the domain Ω that contributes to C_0^t .

3.2.1.2 Estimating the State Density

The state density $q(d)$ reflects the abundance of distance d in a given cellular context defined by the cell boundary $\partial\Omega$ and the objects Y contained therein. Each object $\mathbf{x}_i \in X$, interacting with the objects in Y , will end up somewhere in Ω . The distance d_i associated with the location of \mathbf{x} in Ω depends on the interaction and the frequency with which this distance is *present* in Ω . Knowing Ω and Y , $q(d)$ is fully determined. In some cases, however, the locations of Y might not be known explicitly, but only in a statistical sense. For a few stochastic processes generating the locations of the objects in Y analytical expressions for the *expected* $q(d)$ can be found. For example, if M circular objects with constant radius R are uniformly randomly placed in Ω (that is, they are generated by a *Poisson process* that is conditioned by the number M of *events*), the state density is approximately given by [32]:

$$q(d) \approx 2\pi(d - R) \frac{M}{|\Omega|^M} (|\Omega| - \pi(d - R)^2)^{M-1}, \quad (3.5)$$

where $|\Omega|$ denotes the area of the domain Ω . Figure 3.3B illustrates such a situation. The sketched state density roughly corresponds to the shape of the above expression. Note that the above approximation is only valid in two dimensions. Boundary effects are not taken into account, and objects \mathbf{y} may be overlapping. Therefore, the approximation is inaccurate for small domains or high object densities.

Departure of $p(d)$ from the expected state density $q(d)$ may result from two (independent) effects: one is a potential interaction between objects in X and Y ; the other is that the objects in Y are not distributed according to the model used. Usually, the former is of greater interest than the latter, and it may thus be favorable to use all information about Y that is available.

Given knowledge about both Ω and Y , the state density can be determined by a straightforward sampling procedure: Positions \mathbf{x} in Ω are sampled exhaustively on a uniform Cartesian grid with spacing h . h can be smaller than the resolution of the images from which the objects Y were extracted. In following, $h = 0.25$ pixels is used. For each \mathbf{x}_i , the distance d_i to

the nearest neighbor in Y is then computed. Using this finite sample of distances $D = \{d_i\}_i$, an approximation of $q(d)$ can be found by any density estimation technique, for example Gaussian kernel density estimation.

3.2.2 THEORETICAL DISTANCE DISTRIBUTION

As stated in Section 3.2.1.1, information about interactions is only contained in *deviation* of the observed co-localization – or distribution of distances – from a base-level that has to be expected in the absence of interactions. For the overlap measure, this base-level, C_0^t , is the co-localization measure that would be observed under the hypothesis H_0 : “no interaction”. Under no interaction, all objects in X would be distributed in Ω according to a stochastic process that is independent of the objects in Y . In this view, any statistical dependence between the objects in X and Y is a result of an interaction. This allows defining *interaction* as *the collection of all effects that cause significant correlations between the positions of the objects in X and Y .*

How does a certain deviation from the base level C_0^t relate to interactions between the objects, and what deviations can be considered significant? Ideally, an interaction score is independent of the cellular context and reflects variations of the true interaction strength in a monotonous fashion. The first step toward constructing such a score is a precise definition of the term *interaction strength* in the context of an interaction model.

The development of models for the spatial distribution of interacting objects has attracted considerable attention in the study of ecological systems. Such models are closely related to the mathematical theory of *spatial point process analysis* [140, 32, 107]. Spatial point process (SPP) analysis is a standard statistical framework for studying the spatial distribution of interacting objects. SPP analysis is well-known in ecology, for example in forestry, where the two main branches of SPP, namely descriptive statistics and models for pattern-generating processes, are extensively used. In a recent work on endocytosis [26] a variant of *Ripley’s K -function* [32], which is a descriptive SPP statistic, was used to quantify spatial correlations in the intra-cellular localization of endosomes. A related approach has been used to quantify clustering of GLUT4 storage vesicles in rat adipose cells [91].

The present interaction analysis is derived from the *general binary Gibbs process* with a fixed number of objects, which is a standard model for spatial point pattern generation. The central component of the Gibbs process is an effective pair-wise interaction potential $\Phi(\cdot)$. In many applications, including the present, “interaction” is an abstraction of the different effects that collectively cause an observed spatial pattern. Nevertheless, the mathematical form of the Gibbs process relates to physical models of interacting objects. The interaction potential associates an energy level with each pair $\{i, j\}$ of interacting objects. The probability density of the Gibbs process for two sets of interacting objects, X and Y , has the shape of a Boltzmann distribution:

$$p(X, Y) \propto \exp \left(- \sum_{i=1}^N \sum_{j=1}^M \Phi(\mathbf{x}_i, \mathbf{y}_j) \right), \quad (3.6)$$

i.e., configurations with lower energy occur with higher probability. Equation 3.6 implies mutual independence of the objects within the same set X or Y , in agreement with the assumptions formulated in Section 3.2.1. For *nearest-neighbor interactions*, the interaction potential can be defined as:

$$\Phi(\mathbf{x}_i, \mathbf{y}_j) = \begin{cases} \phi(d_i) & \text{if } \mathbf{y}_j \text{ is NN of } \mathbf{x}_i \\ 0 & \text{else} \end{cases}, \quad (3.7)$$

where the function $\phi(d)$ specifies the strength and distance dependence of the interaction.

Let’s assume a cellular context $\{\Omega, Y\}$ is given. The probability density $p(X|\Omega, Y)$ for the potential in Equation 3.7 then only depends on the nearest-neighbor distances d_i . An inner sum over all j , as in Equation 3.6, is then not required. The mutual independence within X allows factorizing $p(X|\Omega, Y)$ into terms that only depend on a single d_i each:

$$p(X|\Omega, Y) = \prod_{i=1}^N p(\mathbf{x}_i|\Omega, Y) \propto \prod_{i=1}^N \exp(-\phi(d_i)), \quad (3.8)$$

where, unlike in Equation 3.6, an explicit dependence of the potential on \mathbf{x}_i is no longer present.

The probability of observing a certain \mathbf{x}_i is proportional to $\exp(-\phi(d_i))$. That is, as a consequence of the definition of the potential (Equation 3.7), the Gibbs density only depends on the distance d_i associated with the location \mathbf{x}_i . The probability of observing a certain d_i , however, also depends on how frequently an arbitrary object is found at *any* location \mathbf{x} that is a distance d_i away from the nearest object in Y . This frequency is given by the state density $q(d)$ as given in Equation 3.4. Straightforward calculations yield:

$$p(d|\Omega, Y) = p(d|q) = Z^{-1}q(d) \exp(-\phi(d)). \quad (3.9)$$

The normalization constant Z (the *partition function*) renders $p(d|q)$ a true probability density function. Z is defined by an integral over all possible distances in $[d_{\min}, d_{\max}]$:

$$Z = \int_{d_{\min}}^{d_{\max}} q(d) \exp(-\phi(d)) dd. \quad (3.10)$$

In the present model, Z can be obtained by one-dimensional numerical integration. This allows easily evaluating $p(d)$ for different model parameters. This property follows directly from the restrictions made to the general Gibbs process, in particular fixing M , considering only nearest-neighbor interactions, and dropping the position dependence of the potential. In the general Gibbs process, parameter estimation is much involved [10, 6] since computation of Z requires solving a high-dimensional integral. In the present framework standard estimation techniques can be used.

So far, no particular shape of the interaction potential $\phi(\cdot)$ has been specified. The interaction potential can be modeled parametrically or non-parametrically. A specific choice constitutes a hypothesis or assumption about the range, strength, and distance dependence of the interaction. In the following parameterization, these three aspects are represented independently:

$$\phi(d) = \epsilon f\left(\frac{d-t}{\sigma}\right). \quad (3.11)$$

The parameter ϵ is the interaction strength, f encodes the functional shape, σ defines the length-scale, and t is a shift along the distance axis of

the interaction potential. Using Equations 3.9 and 3.11, the joint probability density of observations $D = \{d_i\}_i$ can be specified as:

$$p(D|q) = Z^{-N} \prod_{i=1}^N q(d_i) \exp\left(-\epsilon f\left(\frac{d_i - t}{\sigma}\right)\right). \quad (3.12)$$

This is the central class of models used here to extend co-localization analysis to interaction analysis. All parametric interaction models are formulated as specific instances of this class of models.

In order to understand how, for example, the classical overlap measure relates to an interaction process, a corresponding interaction potential needs to be found. Indeed, the assumptions underlying the simple overlap co-localization measure can be formalized in a specific interaction potential. As discussed, the overlap measure only distinguishes two categories of distances ($d < t$ and $d \geq t$; Equation 3.3). This implies a step-function for the shape $f(z)$ of the interaction potential $\phi(d)$:

$$\begin{aligned} \phi^{\text{st}}(d) &= \epsilon f^{\text{st}}(d - t) \quad \text{with} \\ f^{\text{st}}(z) &= \begin{cases} -1 & \text{if } z < 0 \\ 0 & \text{else} \end{cases}. \end{aligned} \quad (3.13)$$

For the step-potential, the parameters t and σ are redundant. Rescaling of distances is therefore prohibited by setting $\sigma = 1$, which renders t the distance threshold. Irrespective of whether the step-potential is physically (or biologically) relevant, it reflects the implicit assumptions underlying the co-localization measure C^t . Having phrased these assumptions in a statistical model makes it possible to investigate how the co-localization measure C^t represents interactions.

Using the integral definition in Equation 3.3, the co-localization measure C^t can be expressed as a function of the interaction strength:

$$C^t = Z^{-1} \left(\exp(\epsilon) \int_{d_{\min}}^t q(d) dd + \int_t^{d_{\max}} q(d) dd \right), \quad (3.14)$$

with

$$Z = \exp(\epsilon) \int_{d_{\min}}^t q(d) dd + \int_t^{d_{\max}} q(d) dd. \quad (3.15)$$

By definition,

$$\int_{d_{\min}}^t q(d) dd = C_0^t \quad \text{and} \quad \int_t^{d_{\max}} q(d) dd = 1 - C_0^t.$$

Inserting into Equation 3.14 and solving for ϵ yields an estimator $\hat{\epsilon}$ for the model interaction strength:

$$\hat{\epsilon} = \hat{\epsilon}(N, q) = \log\left(\frac{C^t}{1 - C^t}\right) - \log\left(\frac{C_0^t}{1 - C_0^t}\right). \quad (3.16)$$

This estimator is identical to the maximum likelihood estimator for ϵ .

The quantity $\hat{\epsilon}$ corrects for the cellular context and, therefore, fulfills the requirement for a valid interaction score. Equation 3.16 relates the purely descriptive co-localization measure C^t to an interaction model between the objects in X and Y . It thus builds a bridge between patterns in the data (the cellular context as summarized in $q(d)$, and the measure C^t) and functional relationships (interactions) between sub-cellular components. Equation 3.16 further shows that the naive solution $C^t - C_0^t$ is a biased and altogether invalid interaction measure.

Whether an observed estimate $\hat{\epsilon}$ is indicative of the actual presence of an interaction, however, has to be addressed using statistical tests as illustrated in the following section. In principle, tests for significant co-localization as caused by an unknown interaction can already be constructed on the basis of C^t . Reference [154] describes a test procedure that determines the distribution of C^t in the absence of any interaction by randomly sampling locations in the cell. The above measure $\hat{\epsilon}$, however, is also useful for quantification. Finally, $\hat{\epsilon}$ is only one instance of a measure for the strength of interaction, which is derived from a specific model potential (Equation 3.13). Similar corrected measures can also be derived from any other model potential.

3.2.3 STATISTICAL TOOLBOX I: HYPOTHESIS TESTING

Based on the model derived so far, the significance of deviations of the observations from the distribution of expected distances in case of no interaction can be assessed.

3.2.3.1 Power Analysis for the Step Potential

In the parameterization of the interaction model (Equations 3.11 and 3.12), the presence of an interaction is equivalent to $\epsilon \neq 0$. Since the estimator $\hat{\epsilon}$ is a function of the random variables in D , it is a random variable itself. Even if the hypothesis H_0 : “no interaction” is true, a non-zero $\hat{\epsilon}$ may therefore occur with finite probability ($\hat{\epsilon} \neq 0$ does not imply $\epsilon \neq 0$). Within certain bounds around 0, a non-zero $\hat{\epsilon}$ is thus not significant. Inference about interactions requires finding a critical estimated interaction strength above which one can reject H_0 on a prescribed significance level α .

This critical interaction strength is determined by the distribution of $\hat{\epsilon}$ under H_0 (null distribution), which depends on the sample size N , q , and the prescribed α . Under H_0 , the product $C^t N$ is binomially distributed with parameters (C_0^t, N) . The binomial distribution follows from the fact that the step potential distinguishes only two groups of distances. The actual value of a distance within one of the groups carries no further information with respect to the interaction strength ϵ . The critical C^t is computed by evaluating the (numerically) inverted cumulative distribution function of the binomial distribution at $1 - \alpha$. The corresponding critical $\hat{\epsilon}$ follows from Equation 3.16.

The dependence of the critical C^t and $\hat{\epsilon}$ on C_0^t and N is shown in Figure 3.4. Since the step potential distinguishes only two groups of distances, C_0^t carries all relevant information about $q(d)$. It can be seen that the minimum significant excess over C_0^t varies only weakly with C_0^t (Figure 3.4A). Obviously, large values of C_0^t in conjunction with small N do not allow rejecting H_0 , even if $C^t = 1$. The critical value of $\hat{\epsilon}$ is highest at the two extremes of C_0^t and lowest for $C_0^t \approx 0.4$ (Figure 3.4B). As for C^t , it can be seen that for large C_0^t and small N , no finite $\hat{\epsilon}$ is sufficiently large to reject H_0 .

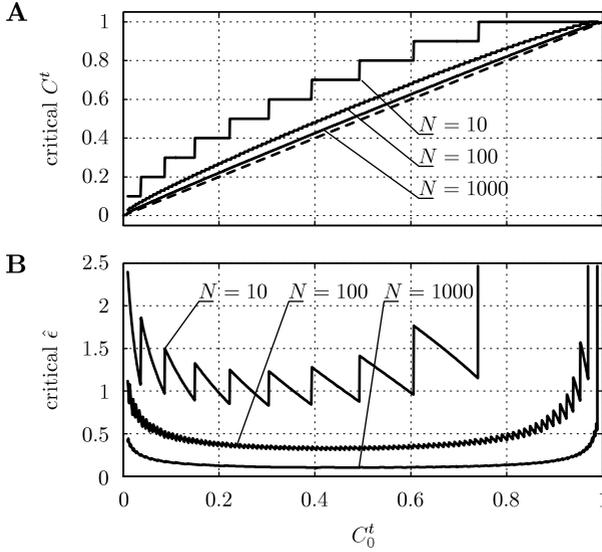


Figure 3.4: Minimum C^t (A) and $\hat{\epsilon}$ (B) that allows rejecting H_0 : “no interaction” ($\alpha = 0.05$) as a function of the base-level C_0^t and different sample sizes N . In A, the expected value of C^t under H_0 is indicated by a dashed line.

The curves in Figure 3.4B show the decision of the statistical test with respect to an estimated interaction strength $\hat{\epsilon}$. A true interaction with strength ϵ greater than this critical value does, however, not guarantee that it will always be detected by the test (type II error: β). Furthermore, a weak interaction may lead to unwanted rejection of H_0 . The behavior of the test critically depends on the *effect size*, which quantifies the departure from H_0 . Here, effect size refers to the true interaction strength $\epsilon = a > 0$. The *statistical power* ($1 - \beta$) quantifies the probability of rejecting H_0 when H_1 : “ $\phi = \phi^{\text{st}}$, $\epsilon = a$ ” is true. Figure 3.5 shows the detection power for an effect size of $a = 1$ as a function of C_0^t . As expected from Figure 3.4B, the power is low at the extremes of C_0^t , eventually dropping significantly below the generally recommended value of 0.8 (dashed line in Figure 3.5), even for $N = 100$. Weak interactions are harder to detect, requiring larger sample sizes to yield a certain power.

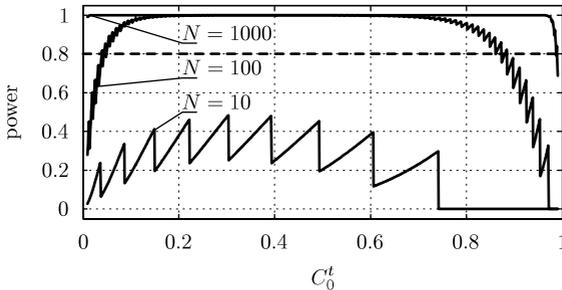


Figure 3.5: Statistical power $(1 - \beta)$ for detecting interactions of a true strength $\epsilon = 1$ as a function of the base-level C_0^t and different sample sizes N .

Achieving high statistical power is a central objective of experimental design, since it directly relates to the success rate of detecting a hypothesized effect. In the design of experimental interaction studies, the robustness and reliability of detecting effects of unknown size should be maximized. Power can be increased by optimizing the experimental design or the subsequent statistical analysis. While increasing sample size might be possible, controlling the cellular context (and hence C_0^t) is not feasible in most practical situations.

Statistical detection power is maximal when all information available in the data is taken into account. The step potential only distinguishes two groups of distances. The precise values of the distances within these groups do not matter for the statistical test. The empirical distance distribution $p(d)$ could deviate strongly from $q(d)$, and yet these deviations might not affect the test statistics $\hat{\epsilon}$ or C^t . In such situations, the statistical detection power may be increased by better modeling the interaction potential, and constructing a more resolving test statistic from this model. This influence of alternative model potentials on statistical power is quantified in the next section.

3.2.3.2 Increasing Power with Non-Step Potentials

Constructing statistical tests as described above requires assuming a specific shape and scale of the interaction potential. In the absence of prior knowledge, however, this model potential can be arbitrarily different from the true potential of the actual biological interactions under observation. Test statistics that are based on a model potential close to the real one can achieve greater power, as will be shown below.

In order to quantify the influence of discrepancies between the model and the true potential, synthetic data consisting of distances drawn from a known distribution $p(d)$ are used. This distribution consists of the state density $q(d)$ and a bias toward certain distances as caused by a known interaction potential $\phi(\cdot)$. A scenario where N objects $\{\mathbf{x}_i\}$ are distributed in a square region Ω containing M randomly placed circular objects $\{\mathbf{y}_i\}$ is considered. The circular objects have identical radii R and are not overlapping. In this case, the state density $q(d)$ is given by Equation 3.5 as shown in Figure 3.6A.

The objects in X interact with the objects in Y according to a *Plummer potential* (with $t = 0$):

$$\begin{aligned} \phi^{\text{Pl}}(d) &= \epsilon f^{\text{Pl}}\left(\frac{d}{\sigma}\right) \quad \text{with} \\ f^{\text{Pl}}(z) &= \begin{cases} -(z^2 + 1)^{-0.5} & \text{if } z > 0 \\ -1 & \text{else} \end{cases} \end{aligned} \tag{3.17}$$

This potential has an overall $1/d$ -shape, but finite value and slope everywhere. The parameter ϵ again controls the interaction strength (potential depth). The parameter σ sets the length scale of the interaction (potential range) and allows gradually changing $\phi(d)$ from a step-like shape to a potential that causes significant attraction toward the objects in Y over large distances (see Figure 3.6B). Consequently, the precise values of measured distances carry information about the interaction potential.

For non-step general potentials, algebraic expressions for $\hat{\epsilon}$ (such as in Equation 3.16 for the step potential) can in general not be derived. In Section 3.2.3.1, the test statistic was based on the fact that the observed

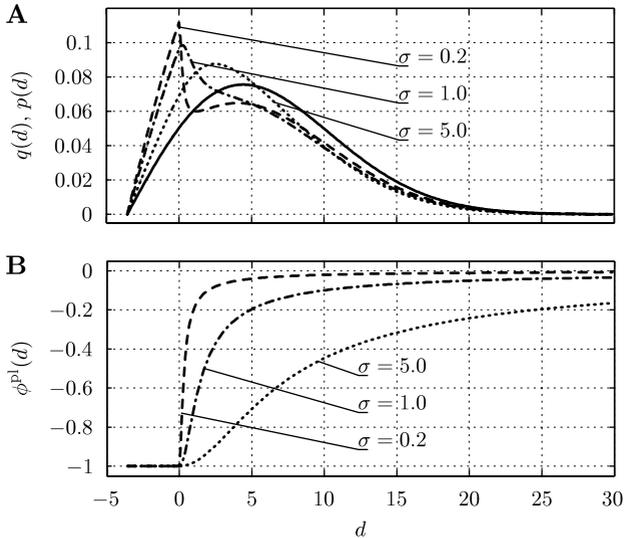


Figure 3.6: Distance distribution for objects interacting according to the Plummer potential. (A) Solid line: state density $q(d)$ for $M = 100$ non-overlapping, circular objects Y with radius $R = 3.57$ randomly placed in a square domain of size 200×200 ; R is chosen to yield a circle-covered area fraction of 0.1. Other lines: resulting distance distribution $p(d)$ for the three potentials shown in B. (B) Plummer potential (Eq. 3.17) with $\epsilon = 1$ and varying scale parameter. Dashed line: $\sigma = 0.2$; dash-dotted line: $\sigma = 1.0$; dotted line: $\sigma = 5.0$.

number of co-localized objects follows a binomial distribution. Now, such reasoning is no longer valid. Statistical tests for the presence of interactions can nevertheless be constructed using different statistics. Since Equation 3.12 describes a member of the exponential family,

$$T = - \sum_{i=1}^N f \left(\frac{d_i - t}{\sigma} \right) \quad (3.18)$$

is a *sufficient test statistic* for ϵ [5]. The concept of sufficient statistics is very powerful. Out of all possible statistics $T = r(D)$ (for any function

$r(\cdot)$ only sufficient statistics carry all information available in D about the unknown strength ϵ of the given potential. For concluding something about ϵ , knowing one sufficient statistic T is thus as good as knowing any other sufficient statistic, or even knowing the entire sample of distances D . A statistic $T = r(D)$ is a sufficient statistic if and only if the joint density of the observations D , $p(D|\epsilon)$, can be factored into two non-negative functions u and v as:

$$p(D|\epsilon) = u(D) \cdot v(T, \epsilon). \quad (3.19)$$

u may depend on the full sample D , but not on ϵ , while v may depend on ϵ , but the dependence on the data must only be through the value of T . Equation 3.12 can be re-written as:

$$p(D|\epsilon) = \left\{ \prod_{i=1}^N q(d_i) \right\} \cdot \left\{ Z(\epsilon)^{-N} \exp \left[-\epsilon \sum_{i=1}^N f \left(\frac{d_i - t}{\sigma} \right) \right] \right\}, \quad (3.20)$$

which proves that Equation 3.18 is a sufficient statistic for ϵ . This is true for any potential parameterized as in Equation 3.11.

For a set of distances D , distributed according to Equation 3.12 with $\phi(d) = \phi^{\text{pl}}(d)$, a test for the presence of interactions can thus be constructed based on $T^{\text{pl}} = -\sum_{i=1}^N f^{\text{pl}}(d_i/\sigma)$ under H_0 : “no interaction”, where the scale parameter σ is assumed to be known. The null distribution of the test statistic can be approximated by i.i.d. Monte-Carlo (MC) samples $\{T_k^{\text{pl}}\}_{k=1}^K$. They are obtained by sampling $N = |D|$ distances d_i from $q(d)$ using the inversion method, computing T_k^{pl} , and repeating this procedure K times.

Given these MC samples, the significance of an observed value of T^{pl} is assessed in a rank-based test. For this, the observed value of T^{pl} is ranked among the $\{T_k^{\text{pl}}\}_{k=1}^K$. If it ranks higher than the $\lceil (1 - \alpha)K \rceil$ -th, H_0 is rejected on the significance level α [5]. This test can be made arbitrarily accurate by increasing the number of MC samples K .

What is the statistical power of this test? Assume H_1 : “ $\phi = \phi^{\text{pl}}$, $\epsilon = a$ ” is true. The statistical power of the rank-based test to reject H_0 when H_1 is true can be estimated with additional MC simulations: For a fixed effect size $a > 0$, one draws N distances d_i from $p(d)$ (again using the inversion

method), computes T^{pl} , and conducts the test as described above [5]. This procedure is repeated many times and the fraction of tests rejected serves as an estimator of the statistical power. Clearly, the power will depend on the effect size a , the number of samples N , and the test statistic used.

In order to quantify the influence of the model potential on statistical power, we test H_0 against H_1 and H_2 : “ $\phi = \phi^{\text{st}}, \epsilon = a$ ” on data generated under H_1 for varying σ (see Figure 3.6B for the true interaction potentials under H_1). The idea behind this benchmark is that one might not know the shape $f(\cdot)$ or the scale σ of the true potential in a practical application. In order to compute the value of a specific sufficient statistic, however, one needs to assume a certain $f(\cdot)$ and σ . The choice might be wrong, and some choices may yield better test performance than others.

Testing H_0 against H_2 uses the sufficient statistic $T^{\text{st}} = -\sum_{i=1}^N f^{\text{st}}(d_i)$, which is proportional to C^t for $t = 0$. As opposed to T^{pl} , this statistic only contains information about the signs of the d_i . The information contained in the precise value of the d_i is ignored. Therefore, this statistic should yield a less powerful test. For the test based on T^{pl} , the true scale σ is assumed to be known (a strategy to estimate the shape and scale of an unknown potential is presented Section 3.2.4.4).

Figure 3.7 shows the number of samples N required to reach on average 80% power as a function of the strength a of the true interaction potential. It can be seen that the power of a test based on the true interaction potential (solid lines) is higher than the power of a test based on a step potential (dashed lines). Moreover, this difference strongly increases with increasing potential range σ : for $\sigma = 5$ the statistic based on the step potential requires 4 times more samples than the sufficient statistic. If the true potential is close to a step potential ($\sigma = 0.2$), both tests perform comparably well. Figure 3.7 also shows that interactions over longer distances are harder to detect. In conclusion, one has to be careful when assuming a step potential (as implicitly done in traditional co-localization analysis), since it may yield low statistical detection power where other model potentials would perform well. Optimal power requires prior knowledge about the true interaction potential. If available, such prior knowledge can easily be included in the present framework by choosing t , σ , and $f(\cdot)$.

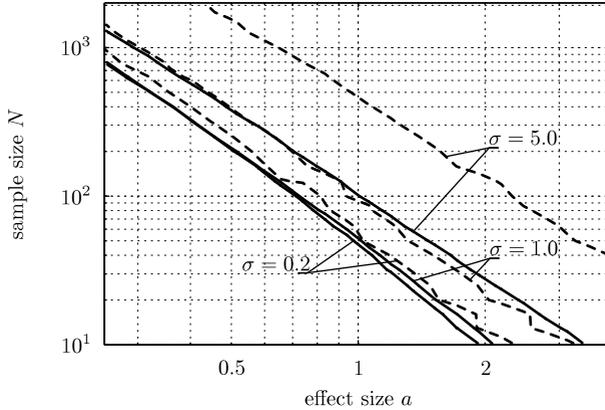


Figure 3.7: Monte-Carlo estimates of 80%-power isolines in the N - a -plane; dashed lines: tests based on T^{st} , solid lines: tests based on T^{pl} . The larger kinks in the dashed lines are due to the discreteness of T^{st} and are statistically significant. Results for the three potentials shown in Figure 3.6 are shown as indicated.

3.2.3.3 Non-Parametric Test for Interaction

In the previous section, test statistics were derived from parametric model interaction potentials. It was shown that using test statistics that are based on a correct assumption about the shape and range of the true interaction are advantageous in terms of statistical detection power. In many applications, however, no prior knowledge about the interaction potential is available. Then, non-parametric tests can be designed that do not require assuming a specific potential.

Following [5], a non-parametric test for interaction can be constructed using the distance counts

$$\mathbf{T} = (T_1, \dots, T_L)^t \quad \text{with} \quad T_l = \sum_{i=1}^N \mathbf{1}(t_l < d_i \leq t_{l+1}) \quad (3.21)$$

in L equi-sized bins defined by $L + 1$ strictly increasing thresholds t_l that

span the entire non-zero range of $q(d)$ for a given cellular context. $\mathbf{1}(\cdot)$ denotes the indicator function. Using these distance counts amounts to implicitly assuming that the potential is a piece-wise constant function. The lower the value of the potential in a given bin, the higher the expected number of counts. H_0 : “no interaction” is equivalent to the potential being zero in all bins, and the expected number of counts is proportional to the integral of $q(d)$ over the bin considered. A deviation from the expected values of counts suggests that the true, but unknown, potential is non-zero in the region spanned by the corresponding bins. Since distance counts in the bins will be anti-correlated (if there are many counts in one bin, there have to be less in others), care must be taken not to over-estimate the significance of the collective deviation of a given set of distance counts from the expected values in the individual bins. As before, a Monte-Carlo sample is used to estimate the joint distribution of the distance counts \mathbf{T} .

First, a Monte-Carlo sample $\{\mathbf{T}_k\}_{k=1}^K$ from the null distribution of \mathbf{T} is obtained by sampling $N = |D|$ distances d_i from $q(d)$ using the inversion method. N refers to the number of observations in D that are subject to the test. Based on the Monte-Carlo distances, \mathbf{T}_k is computed and the procedure is repeated K times. The sample $\{\mathbf{T}_k\}_{k=1}^K$ allows approximating the expectation $\mathbf{E}_0(\mathbf{T})$ and co-variance matrix $\mathbf{Cov}_0(\mathbf{T})$ of the null distribution. The final test statistic U is defined as

$$U = (\mathbf{E}_0(\mathbf{T}) - \mathbf{T})^t \mathbf{Cov}_0(\mathbf{T})^{-1} (\mathbf{E}_0(\mathbf{T}) - \mathbf{T}). \quad (3.22)$$

In a second step, \mathbf{T} and U are computed for the set D of observed distances. As in Section 3.2.3.2, a rank-based test is used. A set $\{U_k\}_{k=1}^K$ obtained from an additional MC sample $\{\mathbf{T}_k\}_{k=1}^K$ is generated as described above. The observed U is then ranked among the $\{U_k\}_{k=1}^K$. If it ranks higher than the $\lceil (1 - \alpha)K \rceil$ -th, H_0 is rejected on the significance level α .

The number of bins L influences the performance of the test. For $L = 2$, the test based on the step potential is recovered. For true potentials that strongly differ from a step-like shape, the test will have low power, as shown in Section 3.2.3.2. Increasing L allows resolving finer details in the structure of the observed distance distribution, and therefore provides the possibility to detect several types of deviations from H_0 . Too large values of L , however, again reduce the statistical power of the test, since the expected number of distances in a given bin will then be very low and only

large deviations from the expectation will be significant (this reasoning is illustrated by the results shown in Figure 3.4).

3.2.4 STATISTICAL TOOLBOX II: ESTIMATION OF POTENTIALS

This section introduces parameter estimation methods for model interaction potentials will be discussed.

3.2.4.1 Maximum-Likelihood Parameter Estimation

For a given potential ϕ , the log-likelihood of its parameters Θ , given the observations D and the cellular context $q(d)$, is the logarithm of the joint probability density (Equation 3.12):

$$\begin{aligned}
 l(\Theta|D, q) &= \log \left(\prod_{i=1}^N p_{\phi}(d_i|q) \right) \\
 &= -N \log(Z(\Theta)) + \sum_{i=1}^N \log(q(d_i)) - \phi(d_i; \Theta).
 \end{aligned}
 \tag{3.23}$$

In *maximum likelihood estimation* one aims at maximizing the likelihood of the unknown parameters, given the data, with respect to the parameters. In practical applications it is often more convenient to maximize the log-likelihood, that is:

$$\hat{\Theta}_{\text{MLE}} = \arg \max_{\Theta} l(\Theta|D, q).
 \tag{3.24}$$

Since the logarithm is a monotonic function, the optima of the likelihood and the log-likelihood coincide. For the present interaction model, no general analytical maximum-likelihood estimator can be found. For the step potential, however, it is possible to show that the maximum-likelihood estimator is identical to the expression in Equation 3.16. In all other cases, numerical optimization techniques need to be used. These can be stochastic (e.g., sampling-based) [106, 134] or deterministic (e.g., gradient-based) [15, 111] strategies. The latter require that the interaction potential ϕ can be differentiated with respect to its parameters. In any case, the

normalizing constant Z needs to be computed by numerical integration, because the state density $q(d)$ can have any functional form, and is usually not known in a compact analytical form. As long as the interaction is not too strong, $q(d) \exp(-\phi(d))$ is a well-behaved function and basic numerical integration schemes are sufficiently robust and accurate.

For parameter estimation it is not necessary to evaluate the sum of the log-state density values at the locations of the data ($\sum_i \log(q(d_i))$) in Equation 3.23, since it is not a function of the unknown parameters.

3.2.4.2 Hyper Models for Parameters of the Interaction Potential

The concept of estimating parameters from a single data set can be easily extended to fitting hyper models for the parameters Θ_k on multiple data sets k . Hyper models could, for example, specify the variation of the parameters of the interaction potential with respect to an additional covariate η_k :

$$\Theta_k = g(\eta_k, \Psi). \quad (3.25)$$

This means that the parameters Θ_k are not independent unknowns, but depend on the covariate η_k through the function $g(\cdot)$. $g(\cdot)$ is known up to the hyper parameters Ψ , which have to be estimated from data. Examples of covariates include time, spatial variables, parameters quantifying the strength of experimental perturbations (e.g., the concentration of a chemical), etc.

A simple example considers the estimation of a common scale parameter σ^* of the interactions in a collection of cells, each of which having a different, unknown interaction strength. Here, σ^* is the hyper parameter and $\sigma_k = g(\sigma^*) = \sigma^*$. The hyper model therefore assumes that the interaction processes in all cells acts over the same range with the same shape of the potential. To simplify matters, no covariate is taken into account. Given a collection of N^{cells} cells, with respective cellular contexts $q_k(d)$ and N_k observed distances D_k , the common scale σ^* and the independent

strengths ϵ_k are found by maximizing the pooled log-likelihood:

$$l^*(\{\Theta_k\}|\{D_k\}) = \sum_{k=1}^{N^{\text{cells}}} l(\Theta_k|D_k, q_k) \quad (3.26)$$

with respect to the parameters $\{\Theta_k\} = \{(\epsilon_k, \sigma^*)\}$. The resulting $(N^{\text{cells}} + 1)$ -dimensional estimation problem can efficiently be solved with a nested maximum-likelihood algorithm: In an *outer* optimization, only the scale parameter σ^* is varied. Given a tentative value for σ^* , the parameters ϵ_k are estimated in an *inner* optimization of the corresponding summand of the likelihood in Equation 3.26. The objective function of the outer optimization is the sum of maxima, $\max_{\epsilon_k} l((\epsilon_k, \sigma^*)|D_k, q_k)$.

3.2.4.3 Parameter Identifiability

Maximum-likelihood estimates are asymptotically unbiased and normal. The estimation of model parameters should thus be robust, provided the model parameters are identifiable. Whether this is the case or not crucially depends on the definition of the interaction potential. The Plummer potential, for example, includes several other shapes of interaction potentials as limit cases, such as the step potential for $\sigma \rightarrow 0$ or the flat potential (i.e., no interaction at all) for $\sigma \rightarrow \infty$ (Figure 3.8).

For certain combinations of values σ and ϵ , the parameters may lose their meaning, or become difficult to determine. When σ is large the potential is practically flat, and ϵ can no longer be identified: Varying ϵ shifts the potential vertically, that is, the potential remains unchanged except for an additive constant. Since such a constant is completely absorbed in the normalizing constant Z , the density $p(d)$, and hence the likelihood l , remains unchanged.

Maximum-likelihood estimators are asymptotically normal with a covariance matrix given by the inverse of the observed Fisher information. When parameters are not robustly estimated, this will be reflected in the covariance matrix. For intra-cellular interaction analysis, in particular in the application considering virus–endosome interactions, the number of data

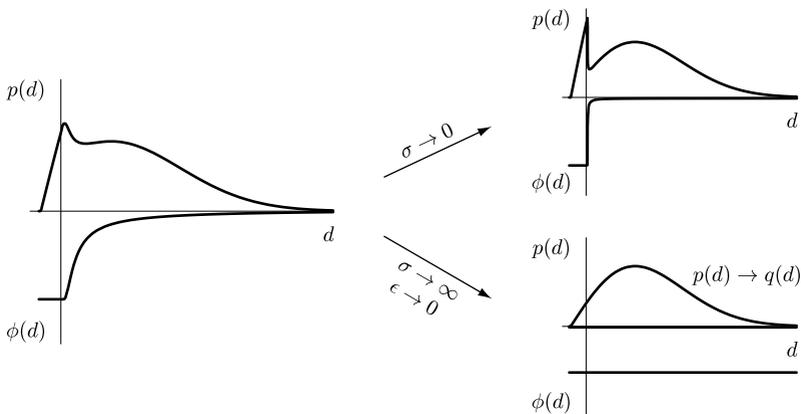


Figure 3.8: Different (limit) shapes of the Plummer potential $\phi^{\text{Pl}}(d)$.

points is typically low (on the order of 100). Normality of the estimator can therefore not be assumed. In the non-asymptotic regime, inspection of the likelihood function around the estimates $\{\hat{\epsilon}, \hat{\sigma}\}$ provides a good visual assessment of the estimation robustness.

Figure 3.9 shows histograms of observed distances along with the fitted distributions $p(d)$ (Equation 3.9) and Plummer potentials for two different cells. The negative log-likelihood function around the estimates $\{\hat{\epsilon}, \hat{\sigma}\}$ is shown below. As a complementary way of assessing estimation robustness, parameter estimates are computed from bootstrap [59] samples of the observed distances: 1000 bootstrap samples consisting of N distances drawn from D with replacement and uniform probability are generated for each of the two data sets. Dots in the likelihood plots depict parameters estimated from the bootstrap samples. In the left example in Figure 3.9, one can see that parameter estimates are scattered in an extended valley of the negative log-likelihood function. This situation corresponds very well to the limit cases of $\epsilon \rightarrow 0$ or $\sigma \rightarrow \infty$ shown in Figure 3.8. In contrast, the estimates in the right example in Figure 3.9 are much more robust. In both cases, however, the model distance distributions fit the observations convincingly. This supports the conclusion that the characteristic shape of the potential, up to a constant, is fitted robustly in both cases. Its

parameters, however, are not always determinable, as illustrated by the left example.

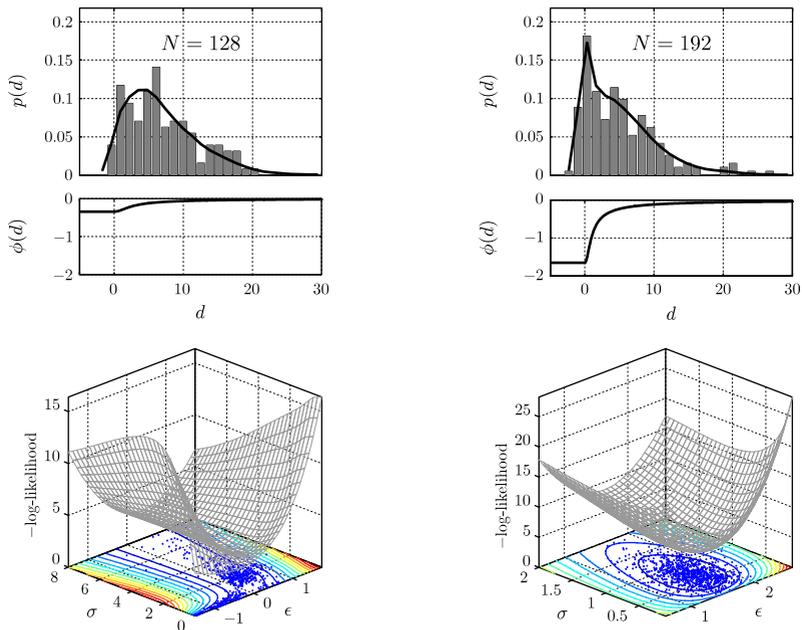


Figure 3.9: Likelihood landscapes (bottom) for the parameters of the Plummer potential, fitted to the distance distributions shown above. Lines in the $\sigma - \epsilon$ -planes depict isolines of the log-likelihood. Blue dots are pairs of parameters estimated from bootstrap samples of the observed distances.

Whether or not parameters are determinable depends on the data: If the distribution of observed distances results from an interaction potential close to one of the limit cases shown in Figure 3.8, the parameters may become meaningless. If the true interaction potential is not close to any of the limit cases, parameters can be robustly estimated.

The present interaction analysis framework allows testing for different potentials. In this *model selection* process, potentials with different shapes and numbers of parameters are fitted independently. The best potential

can then be selected according to, for example, the Akaike or Bayesian information criterion, or a the result of a likelihood ratio test [17]. Furthermore, it may be beneficial to first perform a hypothesis test for the presence of an interaction (see Section 3.2.3), before attempting to fit interaction potentials. Another option is to stabilize parameters, estimation using a prior on the parameters, as discussed in Section 3.2.4.4.

3.2.4.4 Maximum-A-Posteriori Estimators

Maximum-a-posteriori estimators allow stabilizing parameter estimation in case of few data points or poor parameter determinability. The idea is to specify a *prior* $\text{pr}(\Theta)$ on the unknown parameters Θ and maximizing the posterior distribution:

$$\hat{\Theta}_{\text{MAP}} = \arg \max_{\Theta} \frac{p(D|\Theta) \text{pr}(\Theta)}{\int_{\Theta} p(D|\Theta') \text{pr}(\Theta') d\Theta'} = \arg \max_{\Theta} p(D|\Theta) \text{pr}(\Theta). \quad (3.27)$$

Or, equivalently, using the log-likelihood:

$$\hat{\Theta}_{\text{MAP}} = \arg \max_{\Theta} l(\Theta) + \log(\text{pr}(\Theta)). \quad (3.28)$$

For certain models of interaction potentials, the parameter values are naturally bounded. For the Plummer potential, for example, $\sigma \rightarrow 0$ yields a step function. In such a situation, however, the simpler step potential should be used instead. In order to avoid this limit case, a prior on σ can be used to ensure a positive value.

Maximum-a-posteriori estimation can also be used to control the smoothness of non-parametric estimates of the interaction potential. The advantage of non-parametric estimation is that no assumption on the shape of the potential needs to be made. This allows detecting structures in the data that would otherwise be missed (similar to the non-parametric test described in Section 3.2.3.3). The flexibility of non-parametric models, however, needs to be controlled in order to avoid over-fitting. Consider the piece-wise linear (non-parametric) potential $\phi^{\text{n-p}}(d)$ defined as a weighted

sum of kernel functions $\kappa(\cdot)$ centered on P support points d_p :

$$\begin{aligned} \phi^{\text{n.p.}}(d) &= \sum_{p=1}^P w_p \kappa(d - d_p) \text{ with} \\ \kappa(z) &= \begin{cases} |z|/h & \text{if } |z| < h \\ 0 & \text{else} \end{cases}, \end{aligned} \tag{3.29}$$

where $h > 0$ denotes the constant spacing between the support points. Setting $w_P = 0$ enforces that the potential is zero at infinity. Using $\phi = \phi^{\text{n.p.}}$, the remaining weights can be estimated by numerically maximizing the penalized log-likelihood [62]:

$$pl(\Theta|D, q) = l(\Theta|D, q) + \sum_{p=1}^{P-1} \left(\frac{w_p - w_{p+1}}{s} \right)^2 \tag{3.30}$$

with respect to $\Theta = (w_1, \dots, w_{P-1})$. The quadratic penalty in Equation 3.30 corresponds to a Gaussian prior on the differences $\Delta w_p = w_p - w_{p+1}$,

$$\text{pr}(\Delta w_p) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(\Delta w_p)^2}{2s^2}\right), \tag{3.31}$$

which has zero mean and standard deviation s . The smoothness of $\phi^{\text{n.p.}}$ is controlled by the parameter s . Among all potentials of similar global shape, the prior favors the one that has the least oscillations of the weights around the global trend. The larger s , the smoother the estimated potential.

3.2.4.5 Uncertain Distances Measurements

The maximum-likelihood parameter estimates discussed in Section 3.2.4.1 are based on the assumption that the distances $D = \{d_i\}_i$ are known exactly. In any real application, however, these distances are measurements that are corrupted by systematic and random errors. As long as the errors are small compared to the scale on which $p(d)$ changes significantly, they have little effect on the parameter estimates.

Larger measurement errors, however, will lead to a blurring of the ob-

served distribution of distances. Let d' denote the true distance and d the measured distance. Assume the measurement uncertainty $\Delta d = d - d'$ is additive and normally distributed around the true distance d' :

$$\Delta d \sim \mathcal{N}(0, s^2) : \quad p_n(\Delta d) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{\Delta d^2}{2s^2}\right), \quad (3.32)$$

where s^2 denotes the variance of the uncertainty.

Since the true value of any observed distance and its additive uncertainty are independent random variables, the blurring of distances is mathematically described by a convolution of the true density with the density of the measurement uncertainty:

$$p_m(d) = (p * p_n)(d) = \int_{-\infty}^{+\infty} p(\tau) p_n(d - \tau) d\tau, \quad (3.33)$$

where $p(\cdot)$ is the model density for the nearest-neighbor distances. For the model in Equation 3.9), the density $p_m(d)$ of measurements becomes:

$$p_m(d) = \frac{1}{Z\sqrt{2\pi s^2}} \int_{-\infty}^{+\infty} q(\tau) \exp(-\phi(\tau; \Theta)) \exp\left(-\frac{(d - \tau)^2}{2s^2}\right) d\tau, \quad (3.34)$$

and after rearranging:

$$p_m(d) = \frac{1}{Z\sqrt{2\pi s^2}} \int_{-\infty}^{+\infty} \exp\left(\log(q(\tau)) - \phi(\tau; \Theta) - \frac{(d - \tau)^2}{2s^2}\right) d\tau. \quad (3.35)$$

Using this density, a log-likelihood function can be constructed as shown in Section 3.2.4.1:

$$\begin{aligned} l(\Theta|D, q) &= -N \log\left(Z(\Theta)\sqrt{2\pi s^2}\right) \\ &\quad \times \sum_{i=1}^N \log\left(\int_{-\infty}^{+\infty} \exp\left(\log(q(\tau)) - \phi(\tau; \Theta) - \frac{(d_i - \tau)^2}{2s^2}\right) d\tau\right). \end{aligned} \quad (3.36)$$

This means that for *each* distance d_i an integral needs to be computed. Since $q(\cdot)$ is not known in compact analytical form, these integrals need to be

computed numerically. They are approximated with a finite sum over points τ_j on a grid with spacing $\Delta\tau$. For computational efficiency, truncated Gaussians can be used to represent the density of measurement uncertainties. The truncated Gaussians are non-zero only in a local support $\Omega = [-ks, +ks]$, with k sufficiently large. Then:

$$\begin{aligned}
 l(\Theta|D, q) \approx & -N \log \left(Z(\Theta) \sqrt{2\pi s^2} \right) \\
 & \times \sum_{i=1}^N \log \left(\sum_{j \in J_i} \exp \left(\log(q(\tau_j)) - \phi(\tau_j) - \frac{(d_i - \tau_j)^2}{2s^2} \right) \Delta\tau \right),
 \end{aligned}
 \tag{3.37}$$

where the sets J_i contain only those indices j for which $|d_i - \tau_j| < ks$. For maximum-likelihood estimation of model parameters Θ , Equation 3.37 needs to be evaluated many times. Despite the double sum, the likelihood function can be efficiently computed, since the first sum is over few (order 100) data points, and the second sum is only over the few grid points within the finite support of the truncated Gaussians.

3.2.5 APPLICATION OF THE METHOD

The uptake and intra-cellular transport of virus particles is a complex process that involves temporary association with membrane receptors and multiple organelles of the endocytic machinery, such as early and late endosomes, which are the first sorting compartment of clathrin-derived cargo [104]. In many cases, fluorescence microscopy allows resolving the involved entities as discrete objects. This has previously motivated the use of object-based co-localization measures to quantify association kinetics of viruses and endosomes [150] in order to unravel infection pathways. Co-localization analysis focusses on a specific step of endocytic entry program, namely the phase between delivery to the lumen of the endosome and penetration of its membrane. These events, however, are neither instantaneous nor independent of biophysical and biochemical processes, such as vesicle diffusion, active transport, membrane fusion, etc. [102, 51] The observed localization pattern of virus particles and endosomes may thus be influenced by a multitude of effects. A non-step interaction potential may

suggest that the biophysical and biochemical interactions between cellular structures, viruses, and endosomes play an important role in the execution of the entry program.

In the following, the presented framework of interaction analysis is applied to the virus–endosome localization data presented in Section 2.2.4.5, where the classical co-localization measure was used. This application demonstrates the use and workflow of the statistical tools discussed in Sections 3.2.3 and 3.2.4. The benefit of using the present interaction analysis toolbox is three-fold: it allows statistical interpretation of the results, it resolves finer details in the structure of the data, and it is more robust with respect to uncertainties in the input data.

3.2.5.1 Virus–Endosome Distance Data

The data set of Rab5-positive endosomes in Ad2- and TS1-infected HER 911 cells that was already used in Chapter 2 to demonstrate applications of de-convolving active contours is considered. For simplicity, the interaction analysis is demonstrated on distances determined from 2D projections of 3D images. The presented approach, however, is equally applicable in three dimensions without any changes, provided three-dimensional object detection and segmentation is available. Projecting the data into two dimensions alters the estimated potentials (as it also does for any other co-localization measure), since it distorts both the distance data D and the state density $q(d)$.

High-resolution endosome outlines and virus locations were extracted from dual-color fluorescence microscopy images as described in Chapter 2. Near-est-neighbor distances between viruses and endosomes were measured between the virus locations and the closest point on any endosome outline. The state density $q(d)$ was constructed from the endosome outlines by the sampling method described in Section 3.2.1.2 (using a uniform Cartesian grid with spacing $h = 0.25$ pixel). This means that endosomes are identified with the objects in Y , whereas viruses correspond to the objects in X . Computation of $q(d)$ further required determining the cell boundary $\partial\Omega$. An approximate cell mask Ω was found by low-pass filtering and thresholding the 2D projections of the raw endosome images. Viruses and

endosomes outside the cell masks were excluded from the analysis.

3.2.5.2 Non-Parametric Test for Interaction

Like Ad2, TS1 is known to enter the cell by clathrin-mediated endocytosis, but the mutation inhibits escape from endosomes [66, 43]. This should be reflected in a stronger deviation of the empirical distribution of observed distances D from the null distribution $p(d) = q(d)$ than for Ad2. In the present framework, this translates to a non-flat interaction potential between virus centroids and outlines of Rab5-positive endosomes.

Before modeling an interaction potential, H_0 : “ $\phi(d) = 0$ ” is tested against H_1 : “ $\phi(d) \neq 0$ ” for each imaged cell using a the non-parametric statistical test described in Section 3.2.3.3. The number of bins was fixed to $L = 20$, but the bin boundaries were adapted for each cell to span the whole non-zero range of $q(d)$. The results of the tests are summarized in Table 3.1. The fraction of cells for which H_0 has to be rejected is significantly higher for TS1 than for Ad2, irrespective of the significance level and despite the smaller average number of observed virus–endosome distances N . However, Ad2 exhibits significant interaction with endosomes in half of the cells ($\alpha = 0.05$).

	N^{cells}	$p < 0.05$	$p < 0.01$	N
Ad2	135	70 (52%)	25 (19%)	180±50
TS1	139	128 (92%)	100 (72%)	157±59

Table 3.1: Results of non-parametric statistical tests for interaction between viruses and endosomes. First column: number of cells analyzed; second and third columns: number and percentage of cells for which H_0 was rejected on the indicated significance levels; fourth column: mean \pm standard deviation of the observed number of virus particles per cell.

3.2.5.3 Non-Parametric Fit of the Potential

The results of the statistical tests reported in the previous section indicate that the interaction potential is non-flat for many of the cells. They do, however, not permit any conclusions about the shape or strength of the interaction potential, for which, in addition, no prior information is available. It could, for example, be that nothing beyond a temporary association of virus particles with endosomes is observed. This would suggest that the transport of virus particles before and after the association with endosomes is very efficient. Locations of virus particles outside endosomes would then be completely uncorrelated with the endosome locations. Such a fast loss of spatial correlation would be reflected in a step-like interaction potential. Conversely, a non-step potential would suggest that virus and endosome locations are correlated beyond simple co-localization.

The non-parametric estimation procedure for the interaction potential described in Section 3.2.4.4 is applied to obtain a first idea of the potential's strength and distance dependence. This non-parametric potential can then be used as a template for subsequent specification and identification of more specific parametric potentials. Ignoring, for now, possible variability between cells and virus types, all data are pooled and a common non-parametric potential $\phi^{\text{n-p}}(d)$ is estimated. Taking $\phi = \phi^{\text{n-p}}$ (Equation 3.29), a penalized joint log-likelihood is constructed for all cells as:

$$pl(\Theta|\{D_k\}) = \sum_{k=1}^{N^{\text{cells}}} l(\Theta|D_k, q_k) + \sum_{p=1}^{P-1} \left(\frac{w_p - w_{p+1}}{s} \right)^2, \quad (3.38)$$

where $l(\cdot)$ is given by Equation 3.23. $P = 21$ support points d_p , distributed between -5 and 95 pixel with constant spacing $h = 5$ pixel, and a regularization parameter $s = 2$ were used. The weight of the last point, w_P , was fixed to 0. Numerical maximization of Equation 3.38 was done using the *covariance matrix adaptation evolutionary strategy* (CMA-ES) [57, 56], a black-box global optimization method. The estimated $\hat{\phi}^{\text{n-p}}(d)$ is shown in Figure 3.10. Its shape is clearly different from a step function. The value of s had a small effect on global trends of the shape, albeit slightly reducing the slope. The slow increase of the potential suggests that viruses interact

with endosomes over distances of about 10 pixels ($1\ \mu\text{m}$) from their centers.

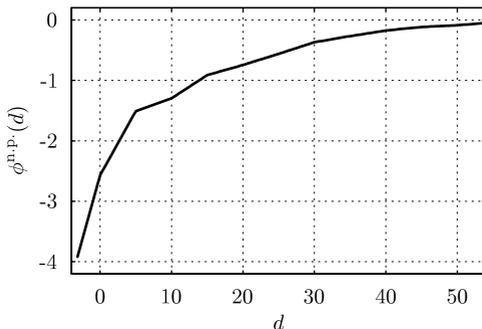


Figure 3.10: The non-parametric estimate of the interaction potential based on all imaged cells.

3.2.5.4 Identification of Parametric Potentials

The most prominent feature of the estimated non-parametric potential is its increase from a single minimum to zero with a characteristic interaction length of about 10 pixels ($1\ \mu\text{m}$). The shape of the potential suggests a $1/d$ -like functional form. However, it must have finite value and slope everywhere.

Parametric potentials are robustly identified from sets of observed distances from individual cells. This allows correlating their parameters with co-variates such as the virus type or the time at which a cell was imaged after infection. The characteristic features of the non-parametric potential motivate a candidate set of four different parametric potentials. Two resemble the shape in Figure 3.10 (*Hermquist* and *Linear type 1*, see Figure 3.11A) and two are generalizations of the step potential with a plateau below $d = 0$ (*Linear type 2* and *Plummer*, see Figure 3.11B). For all potentials, the threshold is fixed at $t = 0$.

All potentials are parameterized as $\phi(d) = \epsilon f((d - t)/\sigma)$ with interaction

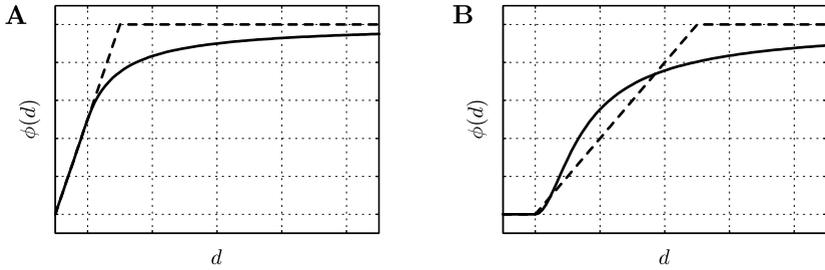


Figure 3.11: Examples of the four non-step potentials. (A) Hermquist (solid line) and Linear type 1 (dashed line); (B) Plummer (solid line) and Linear type 2 (dashed line).

strength ϵ , length scale σ , and threshold $t = 0$. Their shapes $f(\cdot)$ are defined as:

- Hermquist potential:

$$f^{\text{he}}(z) = \begin{cases} -(z+1)^{-1} & \text{if } z > 0 \\ -(1-z) & \text{else} \end{cases} \quad (3.39)$$

- Linear potential, type 1:

$$f^{\text{l1}}(z) = \begin{cases} 0 & \text{if } z > 1 \\ -(1-z) & \text{else} \end{cases} \quad (3.40)$$

- Linear potential, type 2:

$$f^{\text{l2}}(z) = \begin{cases} 0 & \text{if } z > 1 \\ -1 & \text{if } z < 0 \\ -(1-z) & \text{else} \end{cases} \quad (3.41)$$

- Plummer potential:

$$f^{\text{pl}}(z) = \begin{cases} -(z^2+1)^{-0.5} & \text{if } z > 0 \\ -1 & \text{else} \end{cases} \quad (3.42)$$

3.2. OBJECT-BASED INTERACTION ANALYSIS

	$\hat{\sigma}^*$	$\max l^*$	rank
Hermquist	3.96	$-1.2247 \cdot 10^5$	1
Linear type 1	4.14	$-1.2362 \cdot 10^5$	2
Linear type 2	6.61	$-1.2427 \cdot 10^5$	4
Plummer	1.15	$-1.2374 \cdot 10^5$	3
Step	$(t = 0)$	$-1.2632 \cdot 10^5$	5

Table 3.2: Comparison of the estimated common scale parameters of different interaction potentials. The scale parameters $\hat{\sigma}^*$ of potentials as found by maximum-likelihood estimation, and the corresponding maximized pooled log-likelihoods $\max l^*$ (Equation 3.26) are shown for the different potentials.

The potentials can be fitted to the observed distances D_k of individual cells k by maximum-likelihood parameter estimation (Section 3.2.4.1). In order to exclude cell-to-cell variations of the potential scale, the pairs (ϵ_k, σ_k) are not determined for each cell separately. Instead, a single scale parameter σ^* common to all cells is used, while the interaction strengths ϵ_k may vary between cells. Parameter estimation for this model is discussed in Section 3.2.4.2.

The estimated common scale $\hat{\sigma}^*$ and the maximum of the pooled log-likelihood l^* for the four potentials are reported in Table 3.2. As a reference, the values are also given for a step potential with distance threshold $t = 0$. The potentials are ranked according to their log-likelihood. It can be seen that the step potential is outperformed by all others. This remains unchanged even if one compares Akaike or Bayesian information criteria, which take into account the smaller number of free parameters of the step potential. With a difference in log-likelihood of $> 10^3$ to the second-best fit, the Hermquist potential is by far the best fit. It is also subjectively most similar to the non-parametric potential shown in Figure 3.10.

The fits of distance distributions of individual cells are visually convincing, as illustrated by the example in Figure 3.12. The figure shows an example of an imaged cell, infected with TS1, together with the empirical and estimated distance distributions and the corresponding Hermquist potential (images of Ad2-infected cells are visually indistinguishable from those of

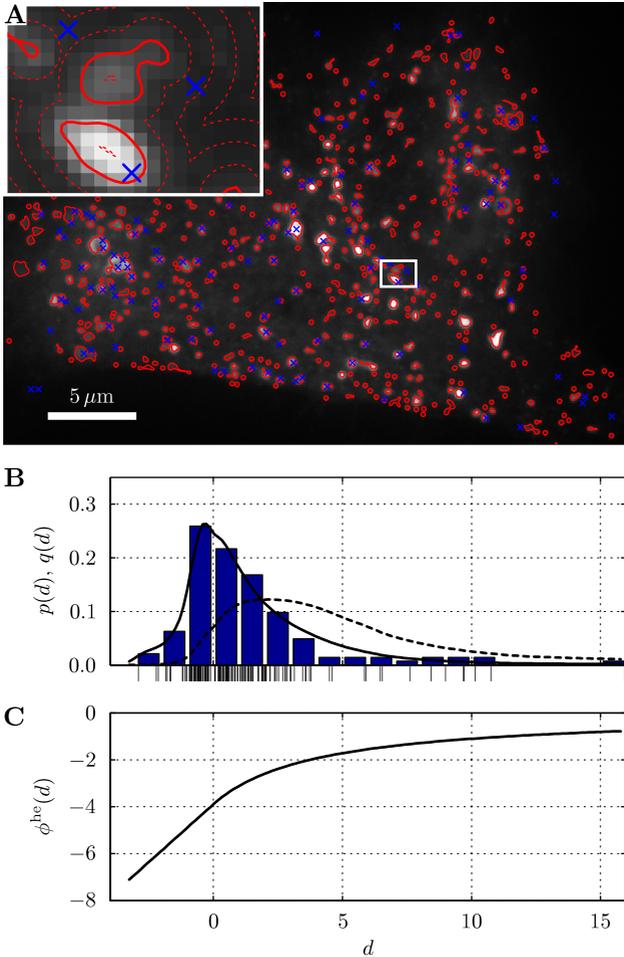


Figure 3.12: (A) Imaged endosomes (Rab5-EGFP) with overlaid outlines (solid red lines) and virus centroid positions (blue crosses). Nearest-endosome-distance isolines (dashed red lines) are shown in the magnified inset. (B) State density $q(d)$ for the shown cell (dashed black line), observed virus-endosome distances (marks and histogram, $N = 143$), and model distance distribution $p(d)$ (solid black line). (C) Corresponding Hermquist potential ($\hat{\epsilon} = 3.90$, $\hat{\sigma}^* = 3.96$).

TS1-infected cells). Despite fitting only one independent parameter (σ^* is fixed from the estimate over all cells), the estimated model distribution captures the features of the data remarkably well. This confirms (at least) that the Hermquist potential is a reasonable choice for the present single-cell data, and not just the best of five bad options. Although the model potentials were motivated by a non-parametric average potential for a collection of data sets, their characteristic shape can be recovered in single cells and across experimental conditions.

3.2.5.5 Covariates

Inspection of the estimated interaction strengths $\hat{\epsilon}$ of the Hermquist potential reveals that it varies within and between the two groups of infected cells. This observation is consistent with the result of the non-parametric test for interaction (Table 3.1), which showed differences between the two virus strains. As expected, these differences are consistently manifested in stronger average interactions for TS1. Within the groups of TS1- and Ad2-infected cells, the strengths vary as well.

The within-group variability comprises statistical fluctuations and natural variations between cells. Since virus internalization and transport is a dynamic process, the time at which a cell was imaged (time post infection) is a further source of in-group variability. Figure 3.13 shows the estimated interaction strengths of a Hermquist potential for all cells infected with Ad2 (crosses) and TS1 (circles) as a function of the time post infection. Throughout the observation period, the interaction strength for TS1 is significantly larger than that for Ad2, confirming the trend reported in Table 3.1. Furthermore, a temporal maximum of the interaction strength is apparent for TS1, while for Ad2 no significant variation over time can be resolved. These results indicate that TS1 and Ad2 use different uptake pathways or exhibit significantly different escape kinetics from Rab5-positive endosomes.

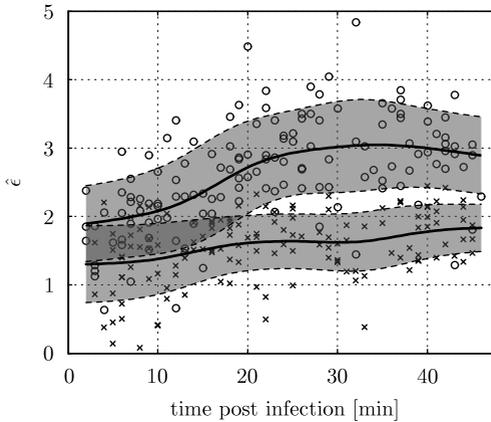


Figure 3.13: Estimated strength of a Hermquist potential (scale $\sigma^* = 3.96$) versus the time post infection. Circles: TS1; crosses: Ad2. The time course of the mean (solid lines) and the ± 1 standard deviation interval (shaded bands) are estimated using a Nadaraya-Watson kernel estimator [132] with a bandwidth of 5 min.

3.2.5.6 Sensitivity to Endosome Segmentation Errors

The present interaction analysis framework is based on the observation of inter-object distances. Imperfect measurement of distances leads to systematic and random errors. In Section 3.2.4.5 it was shown how to include prior knowledge about the distribution of measurement uncertainties. Such prior knowledge about the localization accuracy of virus particles and endosome outlines can, for example, be estimated from benchmarks on synthetic data (see Section 2.2.3.2). Further sources of uncertainty are the systematic over- or under-segmentation of endosomes or the failure to detect an endosome altogether, which has a non-trivial effect on both the measured distances and the state density $q(d)$.

In this section, the robustness of interaction analysis with respect to detection and segmentation errors is tested by artificially corrupting the endosome outline detection and reconstruction. Starting from the endosome outlines Y and cell boundary $\partial\Omega$, high-resolution binary images are cre-

ated that are 1 inside endosomes and 0 outside. The resolution of the binary images was four-fold higher than the resolution of the original image data. Based on these binary images the state density $q(d)$ is estimated as described in Section 3.2.1.2. Virus–endosome distances D are measured as the distance of virus particle locations to the closest non-zero pixel in the binary image. This discretization allows simulating erroneous segmentation and detection by applying the morphological operations of dilation or erosion to the binary image prior to determining $q(d)$ and D .

The effect of erosion and dilation is illustrated in Figure 3.14. While dilation mainly shifts all distances to smaller values, erosion has a more drastic effect whenever it eliminates an entire object. In such situations, the observed virus–endosome distances jump to a much larger value (as seen for the upper-most and lower-most viruses in the left panel of Figure 3.14).

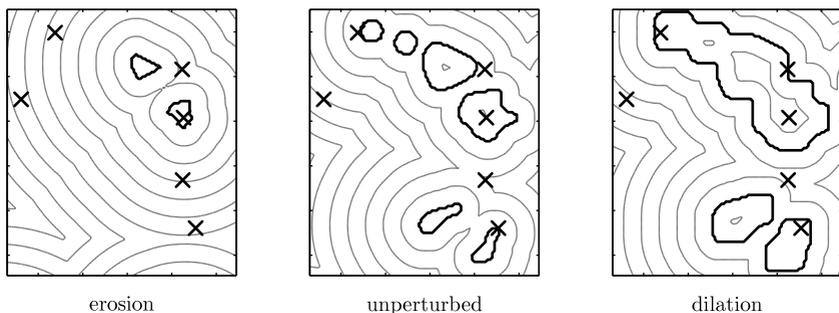


Figure 3.14: Artificially perturbed segmentation and object detection from images. Black lines: segmented endosome outlines; gray lines: distance isolines; crosses: virus locations. Left panel: 4 pixel erosion; middle panel: unperturbed segmentation; right panel: 4 pixel dilation. Units of pixels refer to the upsampled binary image.

Using the distances D and state densities $q(d)$ as estimated from corrupted segmentations, the strength parameter of a Hermquist interaction potential is determined for each cell in the data set. The estimated potential strengths are compared with the classical overlap co-localization measure C^t that is computed using a fixed threshold $t = 0$. In order to account for the increase or decrease of the area covered by endosomes (the esti-

mated cellular context), the simple, but biased (as shown in Section 3.2.2) measure $C^t - C_0^t$ is used. Due to the non-trivial perturbations applied, no attempt is made to model the distribution of measurement uncertainties.

The severity of the perturbations is gradually increased by repeatedly applying the erosion or dilation operation. Figure 3.15 shows the distributions of estimated strengths and corrected co-localization parameters for both virus strains in function of the amount of perturbation applied to the endosome outlines. Erosion and dilation are performed in the binary image up to a distance of 8 pixels in steps of 1 pixel.

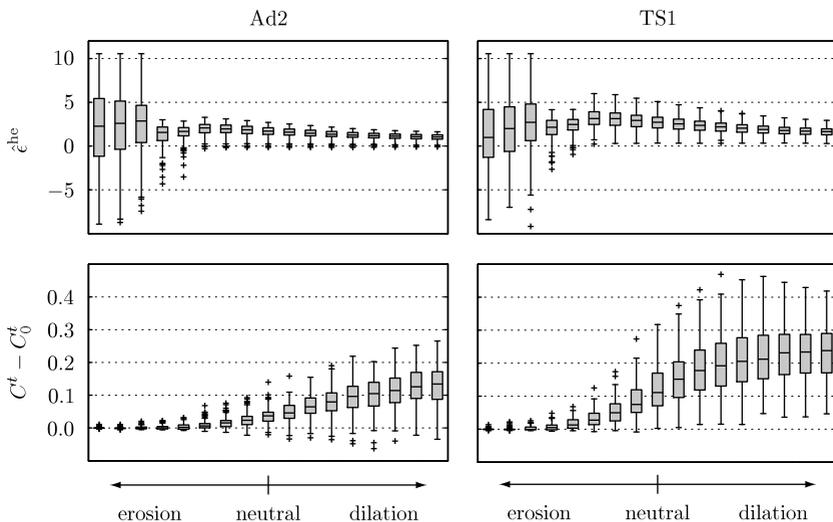


Figure 3.15: Robustness of parameter estimation with respect to segmentation errors. Upper row: Distributions of strengths of Hermquist potential; lower row: distributions of corrected overlap measures. Left column: Ad2; right column: TS1. Edges of boxes are the 25th and 75th percentiles, whiskers extend to the most extreme data not considered as outliers (crosses).

The estimated strengths of the Hermquist potential are fairly robust with respect to the perturbations, except for very strong erosion. The corrected

overlap measure, however, is highly sensitive to the perturbations, both for dilation and erosion. For strong erosion, the corrected measure practically drops to zero, since the area covered by endosomes is reduced to the extent where hardly ever a virus is found inside. For dilation, both the mean and the variance of the corrected measure significantly increase, hence increasing the overlap between the previously well-separated distributions of the two strains. These results illustrate how the present interaction analysis method stabilizes parameter estimation with respect to deficiencies of the image analysis by taking into account the full state density $q(d)$.

3.2.5.7 3D versus 2D

So far, the interactions between viruses and endosomes were estimated from empirical distance data computed from 2D image segmentations. The 2D images were obtained by maximum-projecting the original 3D data sets. Since this projection introduces a bias toward smaller distances in both D and $q(d)$, it likely influences the estimated interaction potentials.

In order to illustrate this influence of the projection, a single cell was analyzed in both 2D and 3D. For reasons of simplicity and comparability between the 2D and 3D data, detection of endosomes and virus particles was limited to finding their intensity centroids (see Figure 3.16 for a 3D visualization). The resulting virus–endosome distances were therefore strictly positive. The state density $q(d)$ was estimated as described in Section 3.2.1.2 for both the 2D and the 3D data.

Using the respective distances D and state densities $q(d)$, Hermquist interaction potentials were estimated for both data sets. Figure 3.17 shows the empirical distance distributions, $q(d)$, $p(d)$, and the estimated interaction potentials. As expected, the projection to 2D increases the frequency of short distances. An even stronger compression effect, however, is also present in the estimate of $q(d)$. The strength of the estimated interaction potential is hence smaller in 2D than in 3D, which suggests that the shortening of distances by the 2D projection is overcompensated by the compression of $q(d)$.

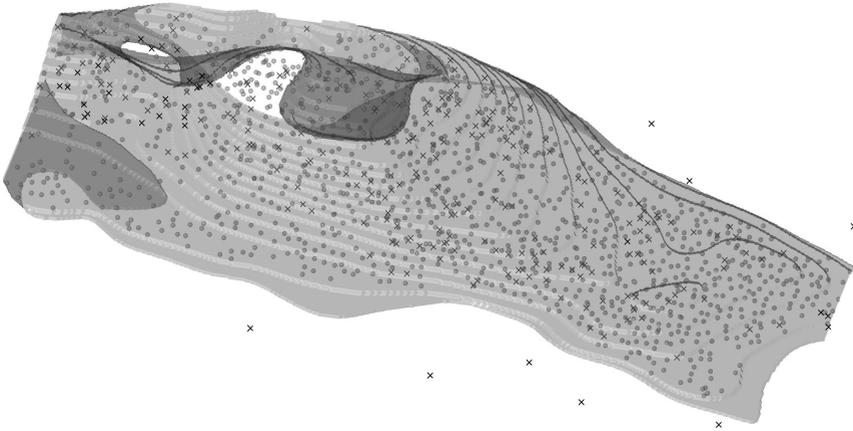


Figure 3.16: 3D rendering of virus particles and endosomes in an HER 911 cell. Shaded surface: cell boundary (darker areas result from invaginations, which are possibly segmentation artifacts); spheres: endosomes; crosses: virus particles.

The effect of projection is most pronounced for intermediate distances. For large distances it is negligible, since they do not occur frequently and mainly reflect distances in the x - y -plane, since the imaged cell is fairly flat. The effect on short distances is also small, since the projection only leads to a minor relative change. This means that, for the presented data, the effect of the projection is strongest where the $3D-q(d)$ is large, while the highest frequency of observed distances is in a region where the projection is less strong. For other data this may, however, be different.

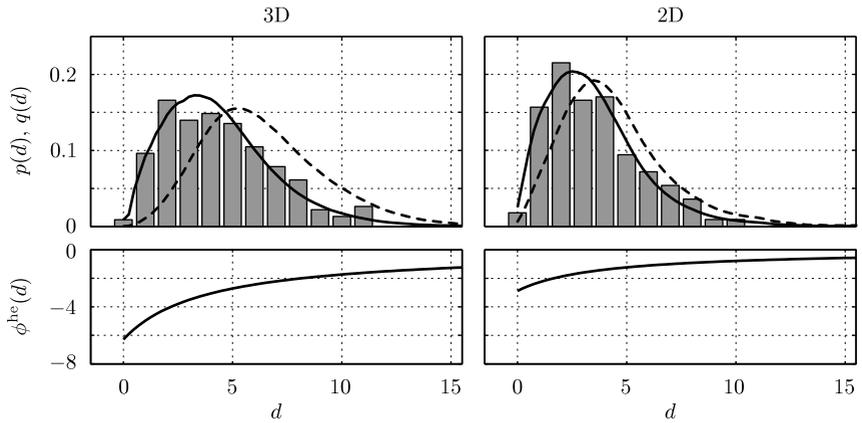


Figure 3.17: 3D versus 2D interaction analysis. Virus–endosome distance distributions (top) and estimated Hermquist interaction potentials (bottom) based on 3D (left, $\hat{\sigma} = 3.80$, $\hat{\epsilon} = 6.30$) and 2D (right, $\hat{\sigma} = 3.72$, $\hat{\epsilon} = 2.89$) image data. Bars: empirical distance distribution; solid lines: $p(d)$; dashed lines: $q(d)$.

MODELING THE MULTI-SCALE TRANSPORT
DYNAMICS OF INTRA-CELLULAR OBJECTS

Cells are complex chemical factories that manage to simultaneously perform countless inter-dependent reactions between huge numbers of reactants. The decryption of the genome and the identification of the proteome of many organisms are rapidly advancing [88]. Based on this vast amount of information, more detailed networks of intra-cellular reactions are being constructed, with the ultimate goal of describing the chemical factory called “cell” as a dynamic system.

The reaction networks, however, are only one part of the story. Reaction networks are typically described in terms of deterministic ordinary differential equations, where the concentrations of the reactants constitute the state variables. This is adequate for abundant molecules in well-stirred systems. Reaction networks with low copy numbers of molecules are usually modeled by stochastic chemical kinetics. In both cases, however, spatial homogeneity is assumed as well in most cases.

This clearly constitutes a strong assumption as the cell is a highly struc-

tured environment. Compartmentalization and spatial concentration gradients are ubiquitous in eukaryotic cells. These physical constraints are an essential component in the regulation of the chemical reaction networks.

Systems of reacting and diffusing species have the potential to generate highly heterogeneous concentration patterns, which are of great importance in developmental biology [145]. The complex internal structure of individual cells, however, largely depends on active – as opposed to passive diffusive – transport mechanisms. While passive diffusion is driven by thermal fluctuations active transport requires energy to move objects around. In cells, active transport is mainly realized by motor proteins that move cargo along the filaments of the cytoskeleton. The integration of physical transport models into spatially resolving models of chemical kinetics is thus one of the great challenges toward understanding how cells work.

Transport processes in cells have been previously described on various levels. The stochastic kinetics, velocity, force generation, and regulation of the activity of single motor proteins have attracted considerable attention [94, 18, 76]. These fascinating molecular machines have the remarkable property that they perform discrete steps along intra-cellular filaments as they transport cargo [76]. Due to their dependence on filaments, volume exclusion effects may reduce the efficiency of transport in cases of high motor or cargo densities. The resulting collective behavior is frequently modeled by variants of the *totally asymmetric exclusion process* (TASEP), taking advantage of the discreteness of the steps [24]. Although most types of motors typically only move in one direction along filaments, bidirectional movement is frequently observed for a great variety of cargos. Whether motors of opposite movement direction simultaneously act on a single cargo, or are selectively switched on and off, is disputed [54]. Regardless of their physical validity, models in which motors of opposite direction work in parallel have reproduced the so-called *saltatory movement* of cargo, and they have been successfully applied to estimate physical parameters of the motor–cargo complex, such as the number of attached motors, or binding and unbinding rates [44].

A further type of model is concerned with the mechanisms behind spatial distributions of intra-cellular objects, such as vesicles and endosomes [75, 34, 33]. Frequently observed patterns such as *aggregation* around the nu-

cleus, *hyper-dispersion* to the periphery of the cell, or *areal* and *radial dispersion* have been reproduced by a single model of organelle transport [34]. In this model, organelles intermittently switch between active transport on filaments and diffusion in the cytosol. The model is parameterized by the rates of switching between the different transport states, the diffusion constant, and a constant velocity on filaments. Organelles are treated as independent objects, and patterns emerging from collective effects can not be reproduced. Fusion, fission, and volume exclusion effects are not taken into account. Furthermore, it is known that organelles do not necessarily move with constant velocities along filaments [86, 108], as assumed in the model.

More detailed models are required if collective behavior, deformation of organelles, or more realistic movement along filaments have to be taken into account. One promising approach is to explicitly resolve the forces acting on and inside organelles. Forces result from the action of attached motors, viscous drag, collisions between organelles, and tension in response to deformation. While such a model is already interesting in itself, its output is also of great value for building higher-level descriptions of collective organelle behavior.

In this chapter, a novel model for the intra-cellular transport of individual objects is described. One focus is on realistically describing the interplay between motor force generation, cargo movement, and motor binding and unbinding events. The model is assembled from simple systems that are amenable to experimental characterization, such as single motor proteins, which have been extensively studied using optical trapping techniques [70]. The model is built bottom-up from elementary mechanical and chemical laws. State transition rules as applied in previous models [44] are not imposed, but rather emerge as limits of physically meaningful parameters. In Section 4.1 the model is specified and an efficient simulation algorithm for it is described. Section 4.2 reports on a large-scale parameter study that reveals many new behaviors and confirms previously described ones. The dependence of the transport dynamics on physical cargo parameters is studied in detail.

4.1 A TUG-OF-WAR TRANSPORT MODEL

Many intra-cellular cargos are transported along microtubules by motor proteins. Microtubules consist of 13 protofilaments that are polymers of α - and β -tubulin dimers. The protofilaments are aligned in parallel in a tubular bundle structure. An important property of microtubules is their polarity. The two ends are called the plus and minus end, respectively. Motor proteins have a dominant direction of movement: Kinesin, for example, moves toward the plus end, while dynein moves toward the minus end [95, 76].

Bidirectional transport along microtubules has been observed in numerous experimental systems involving a large variety of cargos [152]. Since most motor proteins have a single dominant direction of movement [94, 76, 45, 71], this suggests that either motors of opposite polarity are simultaneously present on the cargo, or motors frequently bind and unbind from the cargo in a coordinated way. The latter is not trivial to achieve and has implications that have previously been proven wrong in at least some systems [54]. Furthermore, it has been reported that multiple motors of the same or different kind are simultaneously present on the same cargo [44, 108, 130], resulting in non-trivial velocity and run-length distributions of the cargo.

At first sight, the situation where motors of opposite polarity compete in a *tug of war* seems counterintuitive: *Why would evolution select such an inefficient transport system?* For energetic efficiency, the activity of plus- and minus-end motors should be mutually exclusive. This, however, requires a regulating machinery that senses the activity of the motors and is able to switch groups of motors on or off according to the current demand. Such a machinery may well be present – and some parts have apparently been identified [54] – but a tug-of-war model provides the physical baseline behavior. Any realistic model for a regulated transport mechanism must build on top of a physical transport model. This physical model describes the behavior of single motor proteins and the cargo under external loads. Regulation of motor binding, unbinding, or stepping activity can then be included through non-constant rates that depend on the internal state of the motor-cargo complex and on external stimuli, such as regulatory mechanisms. This added complexity can be justified if the simpler,

non-regulated model fails to explain observed transport properties, or if one has evidence for its existence.

Tug-of-war models can produce a large variety of transport properties that depend on a number of unknown parameters. Many interesting transport behaviors of tug-of-war models have been reported, but not all are easy to understand [74, 108]. Many effects are related to the non-trivial coupling between individual motors in a tug of war: Motor proteins move in discrete nano-scale steps. The stepping rate, but also the rate of unbinding from a filament, depend on the current load, which in turn depends on the state of all motors in the entire motor-cargo complex (see Figure 4.1). Depending on the types of motors present, their number, and the drag forces acting on the cargo, the transport properties change drastically [74, 130, 133]. It is therefore imperative to describe the different possible behaviors of a realistic tug of war in the greatest possible detail before one can falsify such models on the grounds of counter-intuitive observed transport properties.

Transport properties of tug-of-war models have been derived mathematically [74, 9] and have been obtained from numerical simulations and experiments [108, 44, 133]. The kinetic model of Gazzola and co-workers fitted experimental data remarkably well [44]. Müller and co-workers reported a more refined model that is based on a force balance between all active motors [108]. A similar approach was followed by Soppina and co-workers [133]. All of these models are, however, only valid within certain limits, or they require assumptions about uncertain properties of the motor and the cargo. These assumptions are manifested in simple rules for the cargo state transitions. A detailed model for uni-directional transport has been proposed by Kunwar and co-workers [81]. This model resolves the mechano-chemical cycle of elastically coupled kinesin motors down to the level of ATP binding and hydrolysis.

In the following sections, I will describe a force-based model that requires few assumptions and resolves the dynamics of motors and cargo in detail. Certain previous models can be identified as limit cases of the present model for physically meaningful parameters. The present model reproduces previously reported behavior, generates transport patterns that were so far unknown, and helps in understanding the role of the non-trivial coupling between individual motors engaged in a tug of war. The model is

assembled from well-characterized components and can be efficiently simulated. Its structure and the associated numerical simulation algorithm permit many extensions that allow studying more complex systems than the ones considered in Section 4.2.

During write-up of this thesis, the model presented in [81] was published in a simplified form [80] for the limit of saturating ATP concentration. In this limit, the used state transition rules reduce to the model presented in this chapter, which is then simulated using an approximate Monte Carlo scheme. The study in [80], however, focusses mainly on the case of uni-directional motion, whereas the present study is concerned with characterizing the transport properties of a symmetric tug of war.

4.1.1 SPECIFICATION OF THE MODEL

A single, rigid cargo object is moved by several motors of opposite movement direction. Motor proteins stochastically step on, bind to, and unbind from filaments with rates that may depend on the force acting on the motor–filament bond. The physical links between motors and cargo may play a key role in the cargo dynamics, as they define the magnitude of forces and their fluctuations as a response to stochastic events. In the present model, the positions of the motors and the cargo, as well as the resulting forces, are explicitly resolved. An overview of the modeled state variables, their relations, and state transitions is given in Figure 4.1.

Individual motors are coupled indirectly via the cargo. The cargo–motor connections are modeled as linear springs, parameterized by their elasticity κ . This allows linking the positions of motors relative to the cargo with the forces acting between the motors and the cargo. In a response to the motor forces, the cargo moves in the viscous environment of the cytoplasm. Since motors do not “see” each other, transmission of information between motors exclusively takes place via the movement of the cargo. Cargo movement is governed by Stoke’s law of drag, parameterized by the drag coefficient γ . Inertial forces (small Reynolds number) and thermal fluctuations of the cargo position are neglected. The latter is justified because the mechanical energy released by motor proteins is one order of magnitude larger than the average thermal energy of the cargo [64].

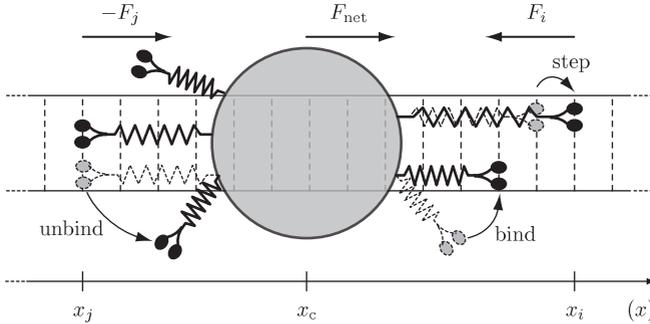


Figure 4.1: Illustration of the tug-of-war transport model. Molecular motors (black things) are attached to a rigid object (gray circle) by elastic springs (zig-zag lines). Motors stochastically bind to, unbind from, and step along discrete binding sites (dashed lines) on a microtubule. The cargo moves in response to the net motor force F_{net} .

4.1.1.1 Time Scales

Molecular motors transform the chemical energy released by ATP hydrolysis to work in a mechano-chemical cycle [73, 18]. In the motor kinesin, for example, this is achieved by a conformational change of the protein structure, which pulls the motor along the filament [119, 21]. The time to complete one such “power stroke” is much shorter than the time between subsequent steps [21]. Stepping of motors can therefore be modeled as stochastic events at discrete times that instantaneously bring the motor-cargo connection into a new state, that is, the motors jump to the next binding site in their movement direction. Binding and unbinding of motors are chemical reactions that are modeled analogously to the stepping events as instantaneous stochastic state changes. The motion of the cargo, however, has to be resolved to greater detail: Depending on the properties of the motors and the cargo, the time between single motor steps and the viscous relaxation time of the cargo position can be of similar order. This separation of time scales into instantaneous stochastic events with small rates and slow continuous movement of the cargo allows building a mathematical model from well-known and well-characterized parts.

4.1.1.2 Mathematical Description

The motion of motors along microtubules is inherently discrete, since the motor–filament interaction that establishes a stable bond is localized to specific parts of the motor and the tubulin dimer. Motors typically do not switch between protofilaments. Step lengths of motors are thus integer multiples of the 8 nm distance between two dimers of α and β tubulin on the same protofilament. Kinesin rarely makes steps of more than 8 nm, while dynein has been observed to also make steps of 16 nm. This motivates the use of $L = 8$ nm as the unit of length.

Each motor i is described by the tuple $m_i = \langle x_i, b_i, d_i \rangle$, where $x_i \in \mathbb{Z}$ is its position, $b_i \in \{\text{bound}, \text{unbound}\}$ its binding state, and the constant $d_i \in \{-1, +1\}$ its movement direction along the filament (x coordinate). The cargo is described by the position of its center of mass $x_c \in \mathbb{R}$. N is the total number of motors bound to the cargo, which is assumed constant. This standard assumption is equivalent to assuming fast binding and slow unbinding of motors to the cargo from an unlimited reservoir of freely diffusing motors [44].

The tension in each motor–cargo link depends on the relative position of the bound motors and the cargo. According to the linear spring model, a force

$$F_i(t) = \kappa(x_i(t) - x_c(t)) \tag{4.1}$$

acts on the motor i and the cargo, provided $b_i(t) = \text{bound}$. For unbound motors the force F_i is zero. The dynamic state changes of the motor–cargo system crucially depend on these forces. In the following, two fundamentally different mathematical descriptions of the dynamics of the cargo and the motors will be used.

During the time τ after the last discrete state change of the motors the cargo position evolves according to an ordinary deterministic differential equation:

$$\dot{x}_c(\tau) = \frac{F_{\text{net}}}{\gamma} = \frac{\sum_{i=1}^N F_i}{\gamma}, \quad x_c(\tau = 0) = x_c^0. \tag{4.2}$$

This equation describes a dynamic balance of forces, that is, the forces exerted by the motors on the cargo are balanced by the viscous drag,

parameterized by the drag coefficient γ . The velocity of the cargo responds instantaneously to changes in motor forces, i.e., there is no inertia. The cell cytoplasm is a non-Newtonian fluid in which the drag coefficient γ may depend on the radius and even the velocity of the cargo in a non-trivial way. For simplicity, it is here assumed that the law of viscous drag for a given cargo–fluid combination may be parameterized by a single constant γ . Using Equation 4.1, the solution of Equation 4.2 is given by:

$$x_c(\tau) = (x_c^0 - \langle x_i \rangle) \exp\left(-\frac{\kappa N_b}{\gamma} \cdot \tau\right) + \langle x_i \rangle, \quad (4.3)$$

where N_b and $\langle x_i \rangle$ denote the number of motors bound to the filament and their mean positions, respectively. The force on a single bound motor i is:

$$F_i(\tau) = \kappa (x_i - \langle x_i \rangle) - \kappa (x_c^0 - \langle x_i \rangle) \exp\left(-\frac{\kappa N_b}{\gamma} \cdot \tau\right). \quad (4.4)$$

The first term on the right-hand side is the force when the cargo is in its equilibrium position. The second term is the transient component of the force, which decays exponentially to zero. Equations 4.3 and 4.4 are only valid between events of binding, unbinding, or stepping of motors, since these events may change the states of motors, their average position $\langle x_i \rangle$, or N . In the following, τ is used as the primary time variable, that is, times are measured relative to the time point of the last event. The real system time t follows by adding up times between events.

Binding, unbinding, and stepping of motors are modeled as inhomogenous Poisson processes. The events are described by the tuples $\mathcal{R}_\mu = \langle \Psi_\mu, i_\mu \rangle$, where the operator Ψ_μ defines the type of the event. Whenever an event μ occurs, the state m_{i_μ} of the associated motor i_μ is instantaneously changed by applying the operator Ψ_μ . The operators are one of $\{\Psi^{\text{on}}, \Psi^{\text{off}}, \Psi^{\text{st}}\}$:

- For **binding events**, that is if $\Psi_\mu = \Psi^{\text{on}}$, the associated motor binds to the filament site closest to the current cargo position: $\Psi_\mu : b_{i_\mu} \leftarrow \text{bound}, x_{i_\mu} \leftarrow]x_c[$.
- For **unbinding events**, that is if $\Psi_\mu = \Psi^{\text{off}}$, the associated motor unbinds from the filament: $\Psi_\mu : b_{i_\mu} \leftarrow \text{unbound}$.
- For **stepping events**, that is if $\Psi_\mu = \Psi^{\text{st}}$, the associated motor

moves to the next filament site in its movement direction: $\Psi_\mu :$
 $x_{i_\mu} \leftarrow x_{i_\mu} + d_{i_\mu}.$

The rates k at which the events occur may depend on the forces in the system, which change continuously in time due to the movement of the cargo:

- **Binding events** occur at a constant rate [74] that *does not* depend on the cargo position, the positions of other motors, or any force in the system:

$$k_\mu = k_{\text{on}}^0. \quad (4.5)$$

This reflects the assumption that a motor that is bound to the cargo, but not to the filament, rapidly explores a space on the filament that contains a large number of free binding sites.

- According to Kramer's law of force-assisted barrier crossing [18, 79] for chemical reactions, **unbinding events** occur with higher rates if the motor–filament bond is under load:

$$k_\mu = k_{\text{off}}(F_{i_\mu}(\tau)) = k_{\text{off}}^0 \exp\left(\frac{|F_{i_\mu}(\tau)|}{F_d}\right). \quad (4.6)$$

This model is well supported by experimental observations [21]. The unbinding force F_d is a motor-specific parameter that sets the force scale; it does not define a maximum force that bound motors can sustain.

- **Stepping events** occur at rates that reflect the empirical or theoretical force–velocity relation of a specific motor type. A motor that makes steps of length L with rate k_{st} moves *on average* with a velocity $v = Lk_{\text{st}}$. For kinesin, for example, the velocity (and therefore the rate of stepping) decreases roughly linearly with external load, that is

$$k_\mu = k_{\text{st}}(F_{i_\mu}(\tau)) = \begin{cases} k_{\text{st}}^0 & \text{if } F \leq 0 \\ k_{\text{st}}^0 \left(1 - \frac{F_{i_\mu}(\tau)}{F_s}\right) & \text{if } 0 < F \leq F_s \\ 0 & \text{else.} \end{cases} \quad (4.7)$$

4.1.2 SIMULATING THE MODEL

The position of the cargo at time τ after the last stochastic event is given by Equation 4.3. Between stochastic events, the rates of the stochastic events are not constant, since they depend on the motor forces. The motor forces (Equation 4.4), however, only vary in function of the cargo position. Therefore, the rates of the stochastic events can be evaluated at any τ .

Formally, the model for the motor dynamics corresponds to a stochastic system of coupled chemical reactions. Simulating the motor dynamics therefore amounts to finding the *next event* and the *next time* at which it happens. The stochastic simulation algorithm (SSA) is the standard method for simulating such systems. In its original formulation, it requires that rates are constant between events. In the present system, however, rates are not constant between events. The mathematical foundation of a variant of SSA that allows simulating such hybrid systems [50] is given in Section 4.1.2.1. A step-by-step specification of the simulation algorithm is presented in Section 4.1.2.2.

4.1.2.1 Hybrid Stochastic-Deterministic Cargo Dynamics

The concept of SSA is to sample from the joint probability density for the next stochastic event μ and the time τ after the last event at which it happens. Say N^+ plus-end and N^- minus-end directed motors are attached to the cargo. In total $3N$ ($N^+ + N^- = N$ binding, stepping, and unbinding events each) different events are distinguished. Each event (except for the binding events) depends on the corresponding motor force.

This joint density for the next event and next time is a time-dependent exponential distribution [46, 50], conditional on the current system state $S = \{m_1, \dots, m_N, x_c\}$:

$$p(\tau, \mu | S) = k_\mu(S, \tau) \exp \left(- \sum_{\mu=1}^{3N} \int_0^\tau k_\mu(S, \tau') d\tau' \right). \quad (4.8)$$

Without any explicit time dependence, finding the time of the next event

amounts to sampling from an exponential distribution, while the next event itself can be selected with a probability proportional to its rate. For hybrid system with time-dependent rates, sampling the time increment is less trivial. It has to be chosen such that

$$\sum_{\mu=1}^{3N} \int_0^{\tau} k_{\mu}(S, \tau') d\tau' = -\log(u), \quad (4.9)$$

where u is a uniform random number in $[0, 1]$. This can be done in two ways: The first option is to evaluate the integral numerically from 0 to increasing τ 's and to stop integration once the equation is fulfilled. The second option is to iteratively solve the equation for τ (for instance using Newton's root-finding algorithm), which, however, requires that the integrands and integrals are known in closed form. If so, the iterative solution of Equation 4.9 should be favored, since it is more efficient.

Here, the linear force–velocity relation allows finding integrals in closed form. For events that can not happen (e.g., binding of an already bound motor) the rates, and hence the integrals, are zero. For all others, the integrals are as follows:

- **Binding event.** k_{on} does not depend on the cargo position x_c , hence:

$$\int_0^{\tau} k_{\text{on}}(S, \tau') d\tau' = k_{\text{on}}^0 \tau. \quad (4.10)$$

- **Unbinding event.** The absolute value in Equation 4.6 requires special attention. Assume the unbinding event concerns motor i . First, one has to check whether the force F_i changes sign between 0 and τ . τ_1 denotes the time of the sign change. If no sign change takes place, $\tau_1 = \tau$. The integral is then split as:

$$\int_0^{\tau} k_{\text{off}}(S, \tau') d\tau' = \int_0^{\tau_1} k_{\text{off}}(S, \tau') d\tau' + \int_{\tau_1}^{\tau} k_{\text{off}}(S, \tau') d\tau'. \quad (4.11)$$

The variable $s = \text{sign}(F_i(0))$ and the movement direction $d = d_i$ of the concerned motor (motor i) are introduced and Equations 4.6 and

4.4 are inserted:

$$\int_0^{\tau_1} k_{\text{off}}(S, \tau') d\tau' = k_{\text{off}}^0 \int_0^{\tau_1} \exp \left[\frac{\kappa ds}{F_d} (x_i - \langle x_i \rangle) - \frac{\kappa ds}{F_d} (x_c^0 - \langle x_i \rangle) \exp \left(-\frac{\kappa N_b}{\gamma} \tau' \right) \right] d\tau'. \quad (4.12)$$

With $c_1 = k_{\text{off}}^0$, $c_2 = \frac{\kappa ds}{F_d} (x_k - \langle x_k \rangle)$, $c_3 = \frac{\kappa ds}{F_d} (x_v^0 - \langle x_k \rangle)$, and $c_4 = \frac{\kappa N}{\gamma}$, this can be written more compactly as:

$$\int_0^{\tau_1} k_{\text{off}}(S, \tau') d\tau' = c_1 \int_0^{\tau_1} \exp [c_2 - c_3 \exp(-c_4 \tau')] d\tau'. \quad (4.13)$$

If $c_3 \neq 0$ (true whenever the cargo is not in its equilibrium position, i.e., it is still moving), then:

$$\int_0^{\tau_1} k_{\text{off}}(S, \tau') d\tau' = \frac{c_1 \exp(c_2)}{c_4} (-\text{Ei}(-c_3 \exp[-c_4 \tau_1]) + \text{Ei}(-c_3)), \quad (4.14)$$

where $\text{Ei}(\cdot)$ denotes the exponential integral function. Note that $\text{Ei}(\cdot)$ is implemented in the “gnu scientific library” (GSL) and can therefore be conveniently evaluated. If $c_3 = 0$, the integral reduces to a simpler form:

$$\int_0^{\tau_1} k_{\text{off}}(S, \tau') d\tau' = c_1 + \exp(c_2) \tau_1. \quad (4.15)$$

The second half of the integral (from τ_1 to τ) is solved in the same way, but with $s = -\text{sign}(F(0))$ and different bounds:

$$\int_{\tau_1}^{\tau} k_{\text{off}}(S, \tau') d\tau' = \frac{c_1 \exp(c_2)}{c_4} (-\text{Ei}(-c_3 \exp[-c_4 \tau]) + \text{Ei}(-c_3 \exp[-c_4 \tau_1])) \quad (4.16)$$

for $c_3 \neq 0$. The simpler result for $c_3 = 0$ is:

$$\int_0^{\tau_1} k_{\text{off}}(S, \tau') d\tau' = c_1 + \exp(c_2)(\tau - \tau_1). \quad (4.17)$$

- **Stepping event.** Due to the piece-wise definition of the force-velocity relations it is necessary to check whether the concerned motor (motor i) stalls ($dF_i > F_s$) or is working at full speed ($dF_i < 0$). Let τ_1 denote the time at which dF_i crosses 0 from below and τ_2 the time at which dF_i crosses F_s from below. The cases when dF_i crosses F_s or 0 from above are analogous, but the bounds of the integrals have to be exchanged accordingly. The integral is split into three parts:

$$\begin{aligned} \int_0^{\tau} k_{\text{st}}(S, \tau') d\tau' &= \int_0^{\tau_1} 0 d\tau' + \int_{\tau_2}^{\tau} k_{\text{st}}^0 d\tau' \\ &+ \int_{\tau_1}^{\tau_2} k_{\text{st}}^0 \left[1 - \frac{d\kappa}{F_s}(x_i - \langle x_i \rangle) \frac{d\kappa}{F_s}(x_c^0 - \langle x_i \rangle) \exp\left(-\frac{\kappa N_b}{\gamma} \tau'\right) \right] d\tau'. \end{aligned} \quad (4.18)$$

Using $c_1 = k_{\text{st}}^0$, $c_2 = 1 - \frac{d\kappa}{F_s}(x_i - \langle x_i \rangle)$, $c_3 = \frac{d\kappa}{F_s}(x_c^0 - \langle x_i \rangle)$, and $c_4 = \frac{\kappa N_b}{\gamma}$ the solution is found as:

$$\begin{aligned} \int_0^{\tau} k_{\text{st}}(S, \tau') d\tau' &= \\ &c_1 \left(c_2(\tau_2 - \tau_1) - \frac{c_3}{c_4} (\exp(-c_4\tau_2) - \exp(-c_4\tau_1)) \right) + c_1(\tau_2 - \tau_1). \end{aligned} \quad (4.19)$$

The times τ_1 and τ_2 at which the force crosses 0 and F_s , respectively, are found by solving Equation 4.4:

$$\tau_1 = \ln \left(\frac{(x_i - \langle x_i \rangle)}{(x_c^0 - \langle x_i \rangle)} \right) \frac{\gamma}{\kappa N_b}, \quad (4.20)$$

$$\tau_2 = \ln \left(\frac{\kappa(x_i - \langle x_i \rangle) - F_s}{\kappa(x_c^0 - \langle x_i \rangle)} \right) \frac{\gamma}{\kappa N_b}. \quad (4.21)$$

τ_1 and τ_2 only exist if the corresponding crossings indeed take place. Because the forces change monotonously between events this can easily be tested by inspecting $F(0)$ and $F(\tau)$ (Equation 4.4).

The calculations above are only valid for the specific force-velocity relation used here (Equation 4.7). For other relations they must be adapted and it may not always be possible to express all integrals in closed form.

4.1.2.2 Algorithm

The time t , the position of the cargo, x_c , and the states m_i of all motors need to be tracked. Starting from an initial state at time $t = 0$, the simulation algorithm cycles through the following steps:

1. Compute the time τ to the next event. This requires solving Equation 4.9) for the upper bound τ of the sum of integrals using Newton's root-finding algorithm. This algorithm solves problems of the type $f(\tau) = 0$. This requires the function value $f(\tau_n)$ and its derivative $f'(\tau_n)$ at a tentative solution τ_n . Here, the function value is

$$f(\tau_n) = \sum_{\mu=1}^{3N} \int_0^{\tau_n} k_{\mu}(S, \tau') d\tau' + \log(u) \quad (4.22)$$

and its derivative is

$$f'(\tau_n) = \sum_{\mu=1}^{3N} k_{\mu}(S, \tau_n). \quad (4.23)$$

Each iteration of Newton's algorithm consists of the following steps:

- (a) For all step events check whether the force acting on the concerned motor crosses 0 or F_s between 0 and τ_n . If so, determine the crossing times τ_1 , and τ_2 using Equations 4.20 and 4.21.
- (b) For each event μ , compute the integrand k_{μ} and integral for the given τ_n . The integral might consist of up to three parts (0 to τ_1 , τ_1 to τ_2 , and τ_2 to τ_n), depending on whether crossings

happened. The integrals are known in closed form for the linear force–velocity relation Equation 4.7 (see Section 4.1.2.1).

- (c) Compute the sums with respect to μ in Equations 4.22 and 4.23.
- (d) Do a Newton step to determine the new tentative τ_{n+1} :

$$\tau_{n+1} = \tau_n - \frac{f(\tau_n)}{f'(\tau_n)}. \quad (4.24)$$

- (e) If τ_{n+1} is not significantly different from τ_n stop the iteration and take $\tau = \tau_{n+1}$ as the solution; else loop back to (a).

2. Update time: $t \leftarrow t + \tau$.
3. Select the next event. This is either a binding, unbinding, or stepping event of one motor:
 - (a) Compute the rates $k_\mu(S, \tau)$ of all events μ (Equations 4.5–4.7).
 - (b) Randomly select the next event. The probability of selecting an event is proportional to its rate. That is, find the smallest μ for which

$$\sum_{i=1}^{\mu} k_i(S, \tau) \geq u \sum_{i=1}^{3N} k_i(S, \tau), \quad (4.25)$$

where u is a uniform random number in $[0, 1]$.

4. Update the cargo position: $x_c^0 \leftarrow x_c(t) \leftarrow x_c(\tau)$ where $x_c(\tau)$ is computed using the system state prior to the latest event in Equation 4.3.
5. Update the state of the motor concerned by the event that happened by applying the corresponding operator Ψ .
6. If $t \geq t_{\max}$ stop; else loop back to 1.

The main complexity of the algorithm lies in step 1, where the standard stochastic simulation algorithm (SSA) can not be used. While the logic is the same the straightforward inversion method for sampling times to the next event can not be applied, because the intensities of the Poisson processes change continuously in time between events. Finding the time to the next event requires on the order of 10 iterations of Newton’s algorithm.

A further complication originates from the piece-wise linear force-velocity relation, which makes crossing detection and case distinctions necessary.

Relative to the other steps, selecting the next event is *not* the computational bottleneck. There is hence no significant benefit in using anything else than the simplest event selection algorithm.

4.2 STUDYING THE MODEL

Some authors [108] have pointed out that the non-trivial coupling between motors leads to a dynamic instability of the balance $N_b^+ - N_b^-$ between bound motors on the plus and minus sides. Even a totally symmetric tug of war (equal numbers of identical motors on each side) can produce long uni-directional motion, but also frequent switching of movement directions. *How can an imbalance of the numbers bound motors on either side be stable, if the system is symmetric?* The reason is that once one side is dominating (say, just by chance), a coordinated sequence of binding events on the opposite side is required to re-establish balance. The time to complete the binding events must be much shorter than the time it takes to move the cargo by an amount that loads newly established motor-filament bonds. If this time is too long, newly bound motors are quickly loaded so strongly that they unbind with a significantly increased rate (Equation 4.6). On the dominating side, however, newly bound motors will more likely remain only moderately loaded, unless, by chance, they outrun the other motors or fail to move at all. The domination by one side is, therefore, self-stabilizing.

Nevertheless, switching of the movement direction does occur. Assuming that two stable imbalances (either side may be dominating) exist, switching amounts to jumping from one stable point to another, that is, the dynamical system needs to cross the barrier between two basins of attraction. This requires, as discussed, a sequence of unlikely events, that is, strong fluctuations around the mean behavior. These events, however, are not statistically independent. They can thus occur with higher probability than a sequence of similar, but independent events. If, for example, the leading motor on the dominating side unbinds, the other motors' load can sharply increase. This increases their unbinding rates, which may trigger

a cascade of further unbinding events and ultimately a switch of direction.

In this view, the magnitude of force changes due to stochastic events is of great importance for the overall system dynamics. In order to allow resolving and adjusting such force fluctuations, the motor–cargo link elasticity κ , and the cargo’s drag coefficient γ , have been introduced in the present model. For elastic links (small κ), single motor steps are of little importance, since the increase in force due to a step is κL . For slow cargos (large γ), the effect of steps is only slowly transmitted between motors. The parameter γ is therefore mainly responsible for the transmission of noise between motors, whereas the parameter κ sets a noise scale.

Increased noise levels have a strong effect on unbinding rates. Since the unbinding rates depends on motor forces through an exponential function (Equation 4.6), the mean unbinding rate changes upon increasing fluctuations of forces. For the same mean force, motors with highly fluctuating forces unbind more frequently. Since the unbinding events are key for reversals of the movement direction, it must be expected that the parameters κ and γ have a strong effect on the transport characteristics of the tug of war. As reported previously [108], tug of wars are also sensitive to motor parameters, such as the unbinding force F_d , the binding rate k_{on}^0 , and the maximum number of motors bound $N^{+(-)}$. While the former two influence the speed with which the system can move away from stable points, the latter relates to the “distance” between the states where one side dominates. The present study mainly focusses on changes in transport properties mediated by the two newly introduced parameters κ and γ . Nevertheless, the maximum number of motors and the binding and unbinding rates will be varied as well, but within narrower bounds and with reduced sampling resolution.

The standard motor model as introduced in Section 4.1.1.2 is used in the following. That is, motors bind, unbind, and step according to Equations 4.5–4.7. The force–velocity relation in this model prohibits back-stepping and keeps rates of forward steps constant under assisting forces.

The motor–cargo system is parameterized by properties of the motors, the cargo, the medium, and the motor–cargo links. In order to reduce the number of free parameters, the symmetric case with equal numbers and physically identical motors moving in the plus and minus direction is studied. Since plus and minus motors and their parameters are very different in reality, this model is not realistic. Nevertheless, the behavior of the symmetric system is a valuable base-line. The symmetric system is expected to show important characteristics of more realistic asymmetric systems, such as long directional runs that are caused by the non-trivial coupling between the motors. Knowing the base-line furthermore allows quantifying the sensitivity of the transport system with respect to asymmetries.

The used motor parameters are comparable to those of kinesin. The rate of forming and breaking motor–filament links (the (un)binding rate) may depend on the cargo and on other factors. The drag coefficient for the cargo depends on the medium and on the cargo size, but not on the velocity. All motor–cargo links have the same elasticity κ , with an upper bound roughly given by the intrinsic elasticity of a single kinesin dimer as determined experimentally [70] and numerically [3]. The effective elasticity, however, can be significantly lowered by deformable cargo, such as vesicles [84]. The unit of length is the length of a single kinesin step, $L = 8$ nm. The unit of force is pico-Newton (pN). Table 4.1 lists the free model parameters and the bounds used in the present study.

In the following sections, these parameters are systematically varied within biologically relevant bounds. Since the main focus is on the effects of κ and γ , they are varied over a larger range with finer resolution (each over 11 logarithmically spaced values within the bound given in Table 4.1). As reported previously [108], k_{on}^0 and k_{off}^0 play an important role for the transport characteristics, as does the maximum number of motors N . In order to reduce the dimensionality of the sampled parameter space, N^+ and N^- only take the values $\{2, 5, 10\}$, which covers the range of values reported for various cargos [54, 130, 108, 44]. The zero-load equilibrium constant $K = k_{\text{on}}^0/k_{\text{off}}^0$ is kept constant at $K = 5$, while k_{on}^0 takes the values $\{0.2 \text{ s}^{-1}, 1.0 \text{ s}^{-1}, 5.0 \text{ s}^{-1}\}$. The average rate of motor binding and the

average number of motors bound are expected to be key determinants of uni-directional run lengths, switching of movement direction, and directed motion velocities. The average rate of motor binding is approximately proportional to the product $N^{+(-)} k_{\text{on}}^0$, while the average number of motors bound is related to $N^{+(-)} K$. k_{off}^0 is thus not varied independently of the other parameters, but is coupled to k_{on}^0 via the constant K .

The stalling force, unbinding force, and stepping rate of motors are fixed to the values reported in [108]. In total, four parameters are thus varied.

	Symbol	Unit	Range
Motor parameters			
binding rate	k_{on}^0	s^{-1}	0.2 – 5.0
unbinding rate	k_{off}^0	s^{-1}	0.04 – 1.0
stepping rate	k_{st}^0	s^{-1}	100
stalling force	F_{s}	pN	6.0
unbinding force	F_{d}	pN	3.0
Cargo parameters			
drag coefficient	γ	$\text{pN s } L^{-1}$	0.004 – 0.667
plus (minus) motors	$N^{+(-)}$	–	2 – 10
Link parameters			
elasticity	κ	$\text{pN } L^{-1}$	0.024 – 4.0

Table 4.1: List of parameters of the present transport model.

4.2.2 AUTO-CORRELATION OF THE CARGO VELOCITY

The main objective of the present study is to robustly characterize qualitative and quantitative changes in cargo motion. Detecting the emergence of stable uni-directional movement is thus central to the analysis. In order to be unbiased, no prior about the length or the velocity of stretches of uni-directional motion should be used.

A simple measure for persistence of uni-directional motion is used: Unless the cargo undergoes a random walk, the velocities from neighboring time points should be correlated. If the cargo moves in the same direction for a long time, the correlation with more distant time points is also high. If the cargo switches direction between two time points, the velocities will be anti-correlated. This motivates using the *auto-correlation function* of the cargo velocity to quantify the persistence of uni-directional motion.

For a given trajectory $x_c(t)$, sampled at discrete times $t_i = i\Delta t$, the auto-correlation function of the velocity $\dot{x}_c(t)$ is given by:

$$R_{\dot{x}_c\dot{x}_c}(j) = (\dot{x}_c \star \dot{x}_c)(j) = \sum_i \dot{x}_c(t_i) \dot{x}_c(t_{i+j}), \quad (4.26)$$

where the sum is taken over all possible i . $j\Delta t$ is the time lag at which the auto correlation is evaluated. A plot of the auto correlation versus time lag is called a *correlogram*. For the present stochastic model, the auto correlation should vanish for large time lags. In fact, a typical correlogram shows an exponential decay of the correlations (see Figure 4.3). The time constant of the decay corresponds to a time scale over which the state of the motor-cargo complex changes significantly, as reflected by a large and persistent change in velocity. For trajectories consisting of phases of processive motion with alternating directions, the time constant of the decay reflects the typical duration of stretches of uni-directional motion between two reversals.

Further time scales, however, may be present in the auto-correlation function. These correspond to other stochastic processes that affect the velocity, such as unbinding and stepping. The auto correlation of the sum of two completely uncorrelated signals is the sum of the two auto correlations of the signals. Stepping events, for example, cause a sharp increase in the velocity, followed by exponential decay. Such jumps are superimposed to the current motion, as long as they do not trigger further stochastic events. Ignoring correlations between the responses to stepping events, unbinding events, and switching of movement direction, this motivates approximating the velocity auto-correlation function with a sum of exponential functions. Each exponential reflects the system's response to a specific type of stochastic event. Based on this, a specific model for the auto-correlation function is introduced in the following section.

4.2.3 MULTI-SCALE DYNAMICS

Cargo trajectories produced by the present model show many characteristics that have been previously described. For suitable parameters, the trajectories show long stretches of fast uni-directional motion with runs directed to the plus and minus ends alternating. Switching of direction is frequently, but not always, accompanied by a deadlock situation in which motors on opposing movement direction neutralize each other. During such deadlocks motion can practically stall but random-walk-like behavior can also be observed. Figure 4.2 shows example time series of the cargo position $x_c(t)$ and its velocity $\dot{x}_c(t)$.

On the time scale of tens of seconds, the most significant feature is the emergence of bi-directional saltatory motion (Figure 4.2A). Phases during which the cargo moves uni-directionally with an apparently constant velocity of about $50 L/s$ typically last 1 to 5 seconds (Figure 4.2B).

On the time scale of seconds, the deadlocks between switches of direction becomes apparent. Phases of directed motion (light gray shading in Figures 4.2C and D) end abruptly with sharp drops in velocity, while fluctuations of the velocity about zero remain during the subsequent deadlocks (dark gray shading). Such deadlocks, however, need not be followed by a switch in movement direction, as seen for example at $t = 31$ s and $t = 32$ s. Figures 4.2C and D further reveal that the velocity during uninterrupted phases of directed motion is not necessarily constant. Between $t = 27.5$ s and $t = 31$ s the step velocities appear to cluster around $-20 Ls^{-1}$ and $-50 Ls^{-1}$, which is reflected in changes of the slope in Figure 4.2C.

The inset in Figure 4.2D, which spans a total of 0.5 s, shows the dynamics of the velocity on the time scale of tenths of seconds. Between instantaneous jumps, the velocity magnitude decays exponentially. Large jumps correspond to unbinding, small jumps to stepping events. The exponential decay is given by the derivative of Equation 4.3. Both stepping and unbinding events essentially change $\langle x_i \rangle$, but the effect of unbinding is typically larger (hence the larger jump). Unbinding events additionally change N_b , thus affecting the time constant of the decay.

In summary, three time scales that can be linked to different stochastic

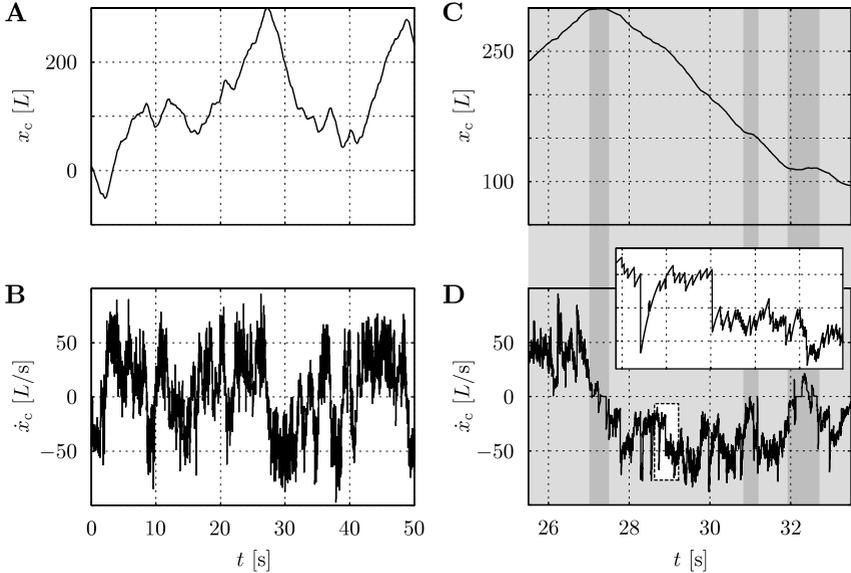


Figure 4.2: Multi-scale dynamics of cargo motion. (A and C) Cargo position; (B and D) Cargo velocity. (C) and (D) are magnifications of the data shown in (A) and (B). The shading separates phases of uni-directional motion (light gray) and mutual blocking of opposing motors (dark gray). (D) illustrates the different time scales for changes of the velocity. The values of free model parameters were: $N^+ = N^- = 5$, $k_{\text{on}}^0 = 1.0 \text{ s}^{-1}$, $\kappa = 1.438 \text{ pN L}^{-1}$, and $\gamma = 0.240 \text{ pN s L}^{-1}$.

events during cargo transport can be identified. Here, the slowest time scale is of highest interest, since it is related to the length of phases of directed motion. As presented in Section 4.2.2, the auto-correlation function of the cargo velocity can be used to quantify the persistence of uni-directional motion.

In order to enable robust estimation of the velocity auto-correlation function, very long ($T_{\text{max}} = 2 \cdot 10^5 \text{ s}$), finely sampled ($\Delta t = 10^{-4} \text{ s}$) trajectories are used. Figure 4.3A shows the velocity auto-correlation function of a trajectory similar to that shown in Figure 4.2 (using the same parameters).

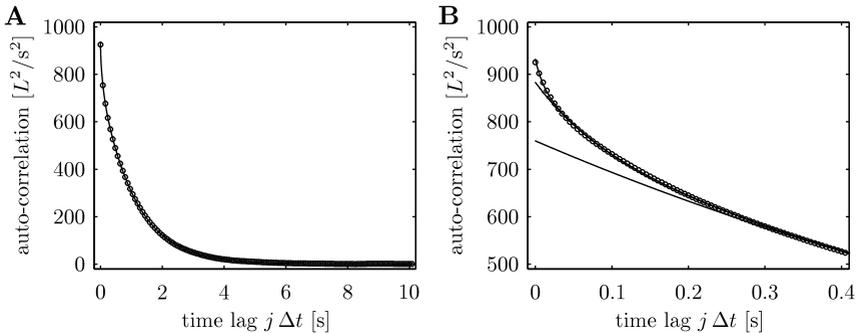


Figure 4.3: (A) Estimated auto-correlation function (circles) and fit (line) with Equation 4.27. The fitted parameters are: $\{a_1, a_2, a_3, a_4\} = \{48.2, 123.3, 758.5, 1.3\}$ and $\{b_1, b_2, b_3\} = \{0.01 \text{ s}, 0.08 \text{ s}, 1.09 \text{ s}\}$. (B) Magnification of the short time-lag region in (A). Upper line: all three exponentials; middle line: fastest time scale left out; lower line: only the slowest time scale displayed. Note that while some time scales are not displayed they were still considered in during fitting.

The estimated auto correlation (circles in Figure 4.3A) appears to decay to zero exponentially, with a time constant on the order of one second. This time constant matches well the typical duration of uni-directional motion as shown in Figure 4.2A. The estimated auto-correlation function is fitted with a mixture of exponential functions. This allows extracting quantitative parameters for comparing the persistence of uni-directional motion in trajectories generated using different model parameters. The estimated auto-correlation function is fitted with a sum of three exponentials, each corresponding to one of the time scales identified in Figure 4.2:

$$R_m(j\Delta t) = \sum_{i=1}^3 a_i \exp\left(-\frac{j\Delta t}{b_i}\right) + a_4 \quad (4.27)$$

with hard constraints

$$0 \text{ s} < b_1 \leq 0.01 \text{ s}, \quad 0.01 \text{ s} < b_2 \leq 0.1 \text{ s}, \quad 0.1 \text{ s} < b_3. \quad (4.28)$$

In good agreement with the initial visual assessment of the sample trajectory, the slowest time scale is fit with a time constant of $b_3 = 1.09$ s. Although the exponential corresponding to the slowest time constant, b_3 , dominates, the other two exponentials and the constant a_4 are necessary for a robust fit. Figure 4.3B shows how successively adding the faster time scales increases the quality of the fit. Ignoring the two fastest time scales would lead to a systematic under-estimation of b_3 , because the increased curvature of the auto-correlation function at small time lags would have to be compensated by an overall larger slope, i.e. smaller b_3 .

In summary, the slowest time scale in variations of the cargo velocity, represented by b_3 , can be used to quantify correlations of the velocity caused by processive uni-directional motion. The influence of the free model parameters on this time scale are studied in the next section.

4.2.4 PERSISTENCE OF UNI-DIRECTIONAL MOTION

The trajectory in Figure 4.2 shows characteristic features that are typically attributed to a tug-of-war transport mechanism: Since in this model motor activity is not regulated, motors of opposing sides can engage in a tug of war that can bring the cargo to a temporary halt. Long phases of uni-directional transport can occur when one side is dominating and a large part of the load is shared between many motors.

Absence of these characteristic motion features, however, is insufficient to reject a tug-of-war model. In the present model, the characteristics of the cargo motion are highly sensitive to parameters of the motors *and* cargo, as shown below.

A robust method for quantifying the persistence of phases of uni-directional motion was demonstrated in Section 4.2.3 for a single set of parameters. Now, the link elasticity κ and the drag coefficient γ are systematically varied within the bounds given in Section 4.2.1 in order to quantify their effect on the velocity auto-correlation function. The remaining parameters are kept the same as in the example shown in Figure 4.2. Figure 4.4 shows the dependence of the slowest time scale, b_3 , on γ and κ .

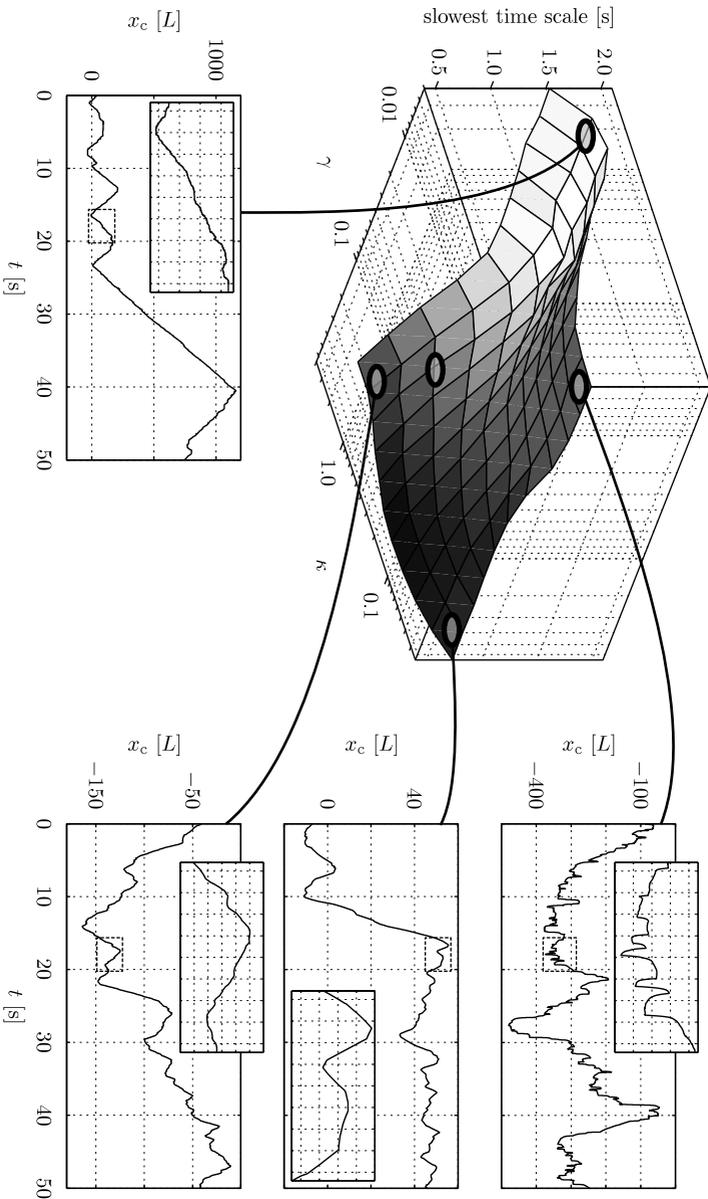


Figure 4.4: Persistence of uni-directional motion as identified by the slowest time scale b_3 of velocity auto correlations. Trajectories are shown for the indicated pairs of κ and γ (circles). In all cases $N^+ = N^- = 5$ and $k_{on}^0 = 1.0 \text{ s}^{-1}$. The circle without inset corresponds to the case shown in Figure 4.2.

b_3 varies between 0.5 s and 2 s, indicating that the persistence of uni-directional motion strongly depends on the properties of the cargo and of the motor–cargo links. These variations in b_3 are accompanied by drastic changes in the visual appearance of the motion on several time scales, as illustrated by four sample trajectories representing the extremes of the parameter space. Interestingly, there are regimes in which b_3 is fairly insensitive to parameter variations, for example on the plateau around small κ and γ .

A clear trend is that b_3 increases with increasing ratio κ/γ . This ratio appears, for example, in the exponential functions in Equations 4.3 and 4.4. For large κ/γ the force of a single motor, and of the motors on the same side of the cargo, rapidly decreases after a step. Forces of opposing motors, however, increase faster. The ratio κ/γ thus determines the speed of *communication* between motors, which helps establish and maintain dominance of one side.

The mobility regimes identified in Figure 4.4 are based on trajectories simulated with $N^+ = N^- = 5$ and $k_{\text{on}}^0 = 1.0 \text{ s}^{-1}$. The number of bound motors and its fluctuation, which is related to k_{on}^0 and k_{off}^0 , have been previously identified as important determinants of transport characteristics [74, 108]. The κ - γ parameter screen shown in Figure 4.4 is therefore extended by also systematically varying the values of $N^{+(-)}$ and k_{on}^0 . Figure 4.5 shows the dependence of b_3 on κ and γ for all combinations of three values of $N^{+(-)}$ and k_{on}^0 .

The most extreme case ($N^+ = N^- = 10$ and $k_{\text{on}}^0 = 0.2 \text{ s}^{-1}$) yields very long phases of persistent uni-directional motion. The dominance of one side over the other is supported by many motors. This redundancy allows tolerating several unbinding events on the dominating side before a deadlock situation can be established. The low unbinding rate additionally makes it unlikely that the sequence of events required for reversing direction happens rapidly. This means that on the side opposite to the current movement direction the load on newly bound motors increases significantly before any further assisting motors can bind.

The trends of the dependence of b_3 on κ and γ , as shown in Figure 4.4, are largely preserved in the plots in Figure 4.5. The magnitude of b_3 , however, varies over two orders of magnitude for the different values of $N^{+(-)}$ and

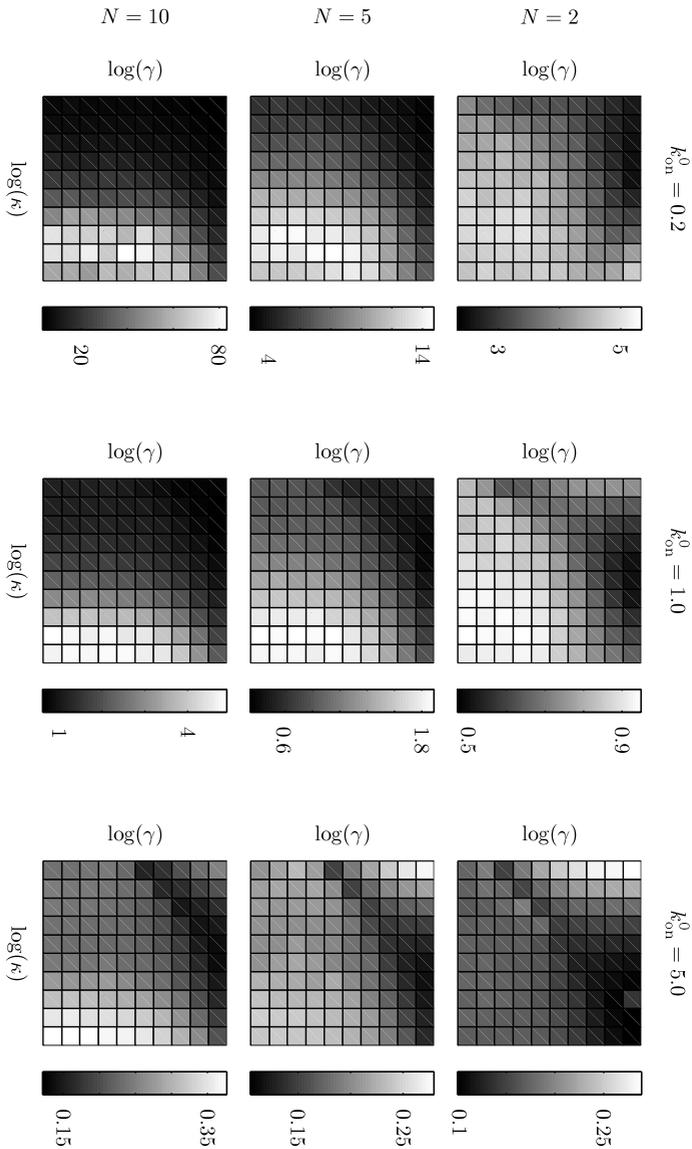


Figure 4.5: Slowest time constant of the velocity auto correlation as a function of all free model parameters. The variation b_g with κ and γ is shown for different values of $N^{+(-)}$ and k_{on}^0 . The same logarithmic scale as in Figure 4.4 is used for κ and γ .

k_{on}^0 used. An interesting detail is that the slowest time scale of the velocity auto-correlation can develop a second peak, which is most easily recognized in the case of $N^{+(-)} = 2$ and $k_{\text{on}}^0 = 5.0$: Beyond a trench oriented along the $\log(\gamma)$ - $\log(\kappa)$ diagonal, b_3 increases for large γ and small κ . This region corresponds to weak coupling between individual motors, as the cargo responds slowly to changes in motor forces. Furthermore, single steps of motors increase the motor's force only slightly. This means that motors can move almost independently of one another and fluctuations of forces are very small. Why this situation *increases* the length of correlations, an effect that was previously attributed to tight coupling between motors, can not be answered based on the data presented so far. In the next section, distributions of cargo velocities are hence analyzed, which may help reveal the cause of this counter-intuitive effect.

4.2.5 VELOCITY DISTRIBUTIONS

In order to further characterize cargo motion, distributions of velocities are determined. As discussed in Section 4.2.3, cargo motion is highly dynamic on multiple time scales. High-frequency components of the velocity, caused by steps and unbinding events of individual motors, are “superimposed” on the overall motion caused by the collective action of all motors. In order to remove the high-frequency fluctuations, time-averaged velocities are determined. Adaptive averaging of the velocity is achieved by computing the net cargo displacement within a time T_v determined from the slowest time scale of the velocity auto-correlation function. The objective of adaptive averaging is to provide velocity distributions that allow identifying phases of deadlocks and uni-directional motion. The time T_v is therefore a tradeoff between averaging out fluctuations and minimizing the blurring of transitions between different phases of motion. $T_v = b_3/2$ was empirically found to be a good compromise. The velocities

$$\langle \dot{x}_c \rangle_i = \frac{x_c(t_i + T_v) - x_c(t_i)}{T_v} \quad (4.29)$$

are computed for all times t_i for which the position of the cargo was sampled. The velocity distribution is determined using a histogram density estimator on a large number of such velocities (i.e., using long trajec-

CHAPTER 4. MODELING THE MULTI-SCALE TRANSPORT DYNAMICS OF INTRA-CELLULAR OBJECTS

ries). Each of the Figures 4.6 to 4.8 shows the velocity distributions for three values of γ and all values of κ . For Figure 4.6 $N^{+(-)} = 2$ and $k_{\text{on}}^0 = 0.2$, for Figure 4.7 $N^{+(-)} = 2$ and $k_{\text{on}}^0 = 5.0$, and for Figure 4.8 $N^{+(-)} = 10$ and $k_{\text{on}}^0 = 0.2$.

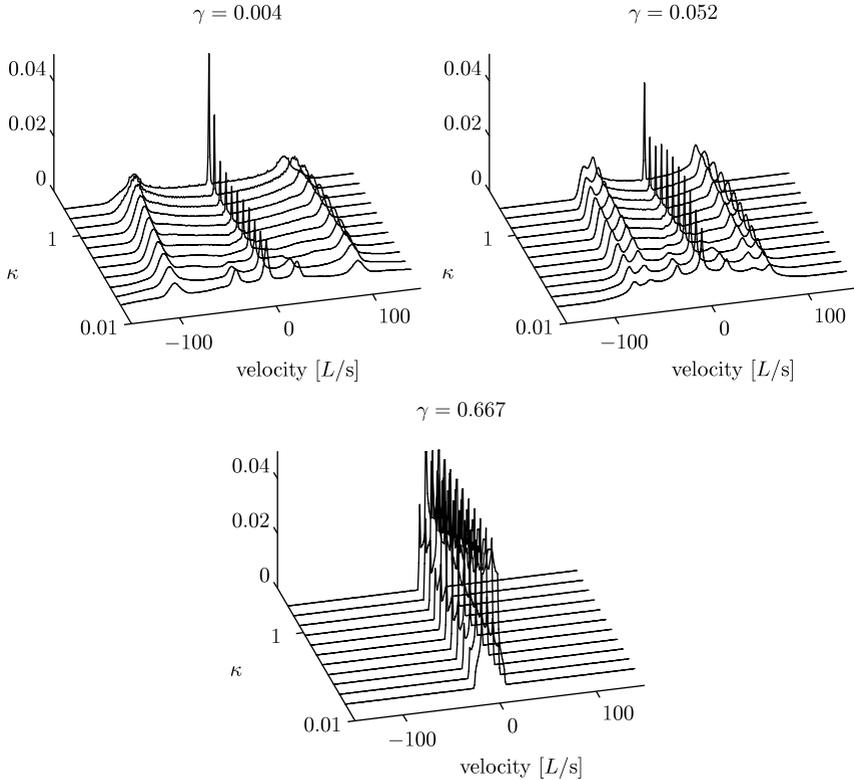


Figure 4.6: Cargo velocity distributions for $N^+ = N^- = 2$, $k_{\text{on}}^0 = 0.2$, and γ and κ as indicated. For better visual appearance, the vertical axis is truncated at 0.04 s/L .

The velocity distributions exhibit significant quantitative and qualitative differences. A trend apparent in all three figures is that velocities decrease with γ . The amount of the decrease, however, also depends on $N^{+(-)}$

and k_{on}^0 . For $\gamma = 0.004 \text{ pNs } L^{-1}$, the drag force exerted by the cargo on the motors is relatively low, which allows them maintaining effective stepping rates (and hence cargo velocities) close to $k_{\text{st}}^0 = 100 \text{ Ls}^{-1}$. For $\gamma = 0.052 \text{ pNs } L^{-1}$, the velocity is reduced, except for when up to $N^{+(-)} = 10$ motors cooperate (Figure 4.8).

Some of the distributions show a pronounced peak around zero, which might correspond to the cargo being in a deadlock situation. In some cases, however, this peak is rather wide, which is inconsistent with a meta-stable deadlock. Alternatively, it could correspond to a situation where the number of bound motors on both sides fluctuates around a common mean value, which could lead to uncorrelated random jumps of the cargo to either side, that is, some sort of diffusion process on the filament. Wide peaks around zero occur for small values of κ and not too small values of γ , which yield weak coupling between motors and small fluctuations of forces upon stepping events. Weak coupling between motors, however, implies that their stepping and unbinding kinetics are largely independent of the overall motor–cargo state. In the limit of completely uncorrelated motors, their actions would have an effect on the cargo that is similar to thermal noise.

A further notable feature of the velocity distributions is the occurrence of peaks at non-zero velocities. For $N^{+(-)} = 10$ (Figure 4.8), for example, two sharp peak appear close to the zero-load velocity of the motors. This indicates that the system mostly is in states where one side dominates and during dominance the state (e.g. number of motors bound) does not vary significantly. Similarly well-resolved peaks are also present in in the distributions shown in Figure 4.6, where, however, the strong peaks at zero remain. Therefore, the system spends considerable amounts of time in tied configurations that do not produce significant net movement.

The distributions shown in Figure 4.6 can have additional peaks appearing close to the peak at zero, the outer peaks, or even close to both. For $\gamma = 0.052 \text{ pNs } L^{-1}$ and small κ , up to seven peaks are clearly resolved. These peaks can be identified with configurations consisting of all possible combinations of 0, 1, or 2 motors bound on either side, with all symmetric configurations mapped to the peak at zero. It has previously been suggested that several peaks in the velocity distribution may occur if the cargo

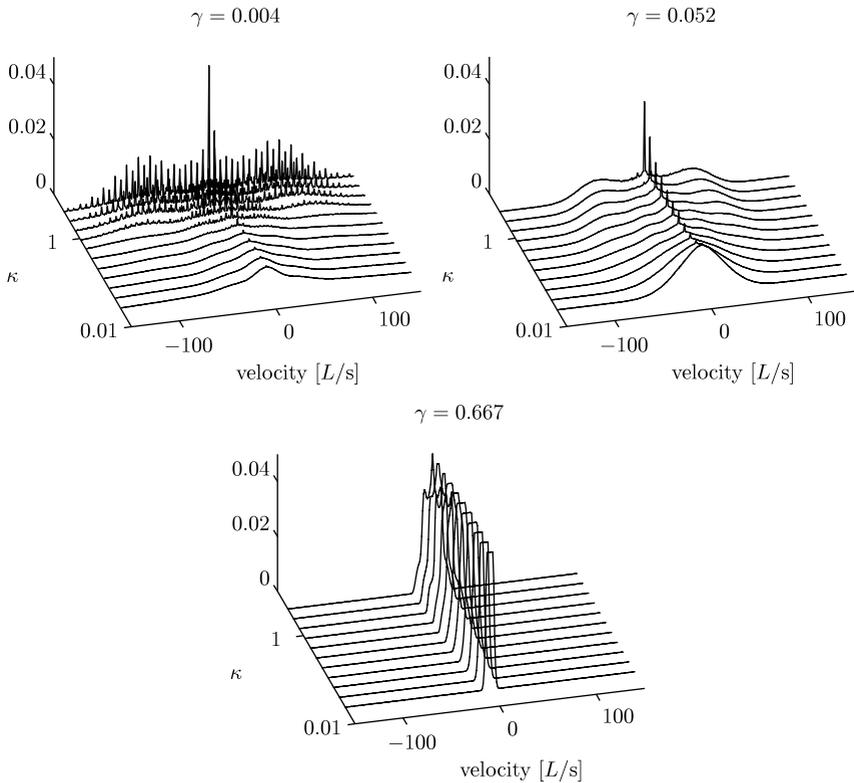


Figure 4.7: Cargo velocity distributions for $N^+ = N^- = 2$, $k_{\text{on}}^0 = 5.0$, and γ and κ as indicated. For better visual appearance, the vertical axis is truncated at 0.04 s/L . Note that the fine peaks for $\gamma = 0.004 \text{ pN s L}^{-1}$ are a quantization artifact, as discussed in the main text.

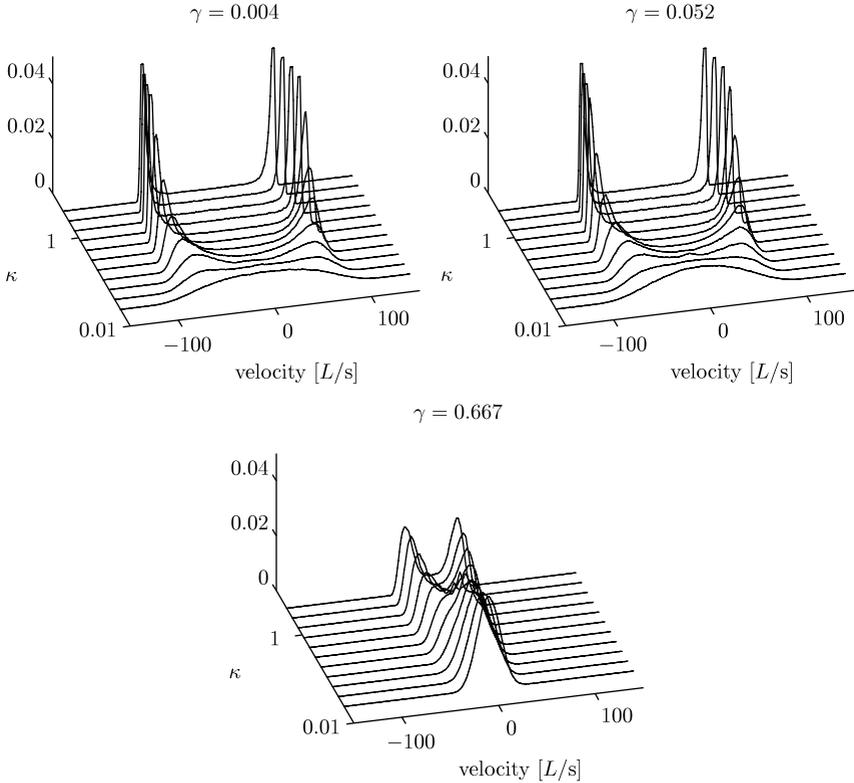


Figure 4.8: Cargo velocity distributions for $N^+ = N^- = 10$, $k_{\text{on}}^0 = 0.2$, and γ and κ as indicated. For better visual appearance, the vertical axis is truncated at 0.04 s/L .

drag force is high [130]. The observation that this phenomenon depends on κ , as shown in Figure 4.6, is new.

For $N^{+(-)} = 2$ and $k_{\text{on}}^0 = 5.0$, no additional peaks are apparent (Figure 4.7), but wide distributions are observed for small γ . In case of $\gamma = 0.052 \text{ pNs } L^{-1}$ and large κ , the two side peaks become apparent.

In Figure 4.7, many sharp peaks occur in case of large κ and small γ . These peaks are not measurement noise, but rather a quantization effect: After a step of any motor, the cargo quickly relaxes to its new equilibrium position, where it comes to a halt. The time scale of relaxation is smaller than the stepping rate of the motors. Computing the velocities with a moving-window average therefore amounts to counting individual steps. This quantization effect becomes apparent if the total number of steps made by all motors during the averaging window is small. In the cases where the peaks are most pronounced, typically as few as 10 steps occurred in a window. For the cases shown in Figure 4.7, the quantization effect is much less pronounced, although the same number of motors was used. The reason is that $k_{\text{on}}^0 = 0.2$. This facilitates maintenance of uni-directional motion supported by two motors (see Figure 4.5). With more motors bound, the number of steps per unit of time increases and the quantization effect is less pronounced.

A noteworthy feature of the cargo velocity distributions is that the maximum velocity does never significantly exceed the zero-load velocity of a single kinesin ($100 L/s$). In contrast to that, higher velocities have occasionally been observed in vivo (see e.g. [44] and Figure 1.10).

The output of the present model seems reasonable, because the mean motor velocity decreases with increasing load. Nevertheless, it has to be taken into account that motors step stochastically. This means that within a finite time a motor's velocity fluctuates around its mean. Therefore, it has to be expected that in a group of motors moving in the same direction some will have advanced further than others. *Could it be that this yields higher net cargo velocities?* In the present model, this is only possible if motors can speed up each other. Three conditions need to be fulfilled for this to happen: First, the stepping rate needs to increase with an assisting load on the respective motor (the force-velocity relation used in the present model does not have this property). Second, leading motors need

to be able to exert assisting forces on rearward motors. In order to allow for motors to be “pushed” forward, the motor–cargo link must not buckle under compression. In the present model it is assumed that springs do not buckle, but there is experimental evidence that kinesin has a significantly lower (if not vanishing) stiffness under compression (see Figure 5 in [70]). The third prerequisite is that already few steps of the leading motor result in assisting loads on the rearward motors. Otherwise, fluctuations would be averaged out and motors would have similar loads. This requires stiff motor–cargo links and fast relaxation of the cargo to its equilibrium position (i.e., low drag forces).

Taken together, the force-velocity relation used here does not allow for groups of motors to speed up each other beyond the single-motor velocity. Furthermore, the resting length of the motor-cargo link is at least the length of the motor protein, which is about $110\text{ nm} \approx 15L$ for kinesin. If pushing forces are excluded due to buckling of the link, an assisting force is only possible if the rearward motor is pulled by the leading one. This requires that they are at least two resting lengths ($\approx 30L$) apart. Taking this into account, it is very unlikely that simple extensions of the present model systematically generate cargo velocities that are higher than single motor–velocities.

The failure to systematically generate higher velocities does, however, not invalidate the present model. Due to their short duration, stretches of fast directed motion on intra-cellular filaments can show higher velocities than the long-term average velocities predicted by the present model. On short time scales, the model can produce cargo velocities that are significantly higher than single-motor velocities due to the stochastic nature of stepping of individual motors.

Intriguingly, it is not easy to imagine how the stepping rate of motors could be increased by other means, e.g. by regulatory molecules: At saturating ATP concentrations, the stepping rate of kinesin is likely limited by the diffusive search of the free head for a new binding site [21]. The rate of this event mainly depends on the structure and elasticity of the neck-linker region, which may not be easily adjusted without altering fundamental properties of the motor (such as the stalling force, which seems to be the same *in vivo* and *in vitro*). In order to explain the full range of

observed cargo velocities, physical effects ignored so far (e.g., head-head interactions) may have to be included in future models.

4.3 RELATIONSHIPS TO OTHER MODELS

The present model includes some transport models used in previous studies as limit cases, which are briefly discussed in the following paragraph. The discussion of these models is followed by a few examples of other related models. Some of these models, however, focus on slightly different systems or use different approaches, which renders a direct comparison to the present model difficult.

The model for the motion of Adenovirus particles on intra-cellular microtubules proposed by Gazzola and co-workers [44] does not explicitly resolve the forces acting on motors and cargo. In their model, steps of the cargo are instantaneous consequences of single motor steps, which occur at constant rate. Each step brings the cargo forward by one unit of length, but has no effect on the binding, unbinding, or stepping rates of any motor in the system. In the present model, this behavior can be recovered by infinitely stiff motor-cargo links ($\kappa \rightarrow \infty$) and zero drag force exerted on the cargo by the surrounding fluid ($\gamma \rightarrow 0$). For virus particles, which are small and rigid objects, these limits seem reasonable. Furthermore, the present force-velocity relation (Equation 4.7) needs to be changed: Each step of any motor needs to be followed instantaneously by steps all other motors. This requires that the rate of forward steps is increased by assisting forces and that back-stepping of motors can be triggered by backward forces. Experimentally determined force-velocity relations of, for example, kinesin, indeed show such phenomena [21].

The tug-of-war transport model of Müller and co-workers [108] is based on equal load sharing between motors. For a given number of bound plus- and minus-end directed motors, a common motor (and cargo) velocity is determined that must be in agreement with the force-velocity relation of all bound motors. The sum of all motor forces has to be equal to the sum of all external forces at all times. In the present model, this amounts to assuming very elastic motor-cargo links ($\kappa \ll 1$). With such links, the motors on either side will be able to step into configurations in which they

all have the same load. Nevertheless, back-stepping of motors is needed to allow for any cargo motion when motors of opposing movement directions are simultaneously bound to the cargo. Since load balance is established instantaneously, $\gamma \rightarrow 0$ is also required.

The model of Kunwar and co-workers [81, 80] resolves forces acting on the cargo that result from stretching of the motor–cargo link and viscous drag on the cargo. In their formulation, the full kinetic cycle of ATP binding and hydrolysis is included. At saturating ATP concentration, the effective binding, stepping, and unbinding rates of kinesin reduce to the ones used in Reference [108] and in the present work. However, Kunwar and co-workers also consider non-linear force-velocity relations of single motors.

Badoual and co-workers [7] presented a model for the collective motion of rigidly coupled motors. The lack of elasticity requires treating the collection of motors as a single object undergoing biased diffusion in a periodic potential. Depending on their state, individual motors “see” a potential that is either flat or has a skewed saw-tooth-like shape. Switching from the saw-tooth potential to the flat one occurs preferentially at the saw-tooth’s minima, hence generating net motion in one direction. In the inelastic limit studied, a separation of time-scales into the power-stroke scale and the cargo-movement scale is not valid. Therefore, their model can not be recovered as a limit of the model presented in this chapter.

Vilfan and co-workers [148] presented a model for motor-driven filament motion as frequently studied in gliding assays. In their model, a linear chain of elastically coupled single-headed, non-processive motors moves on a rigid filament. Individual motors cycle through binding, power-stroke, unbinding, relaxation, and rebinding. The power strokes can move currently unbound motors forward due to the elastic coupling, which yields processive uni-directional motion. Due to the different motors used, this model’s output is incomparable to the one presented in the present chapter.

Campàs and co-workers [19] studied the collective dynamics of molecular motors coupled by an effective interaction potential. In their model, motors move on a lattice with rates that depend on the local arrangement of motors and on external forces. Since the coupling between the motors is extremely strong (rates change instantaneously after motor steps), this model is probably close to the case of very rigid motor-cargo links.

CHAPTER 4. MODELING THE MULTI-SCALE TRANSPORT DYNAMICS OF INTRA-CELLULAR OBJECTS

However, an external drag force on the cargo is not included. Instead, the leading motor bears the total external load, which seems to be an unrealistic assumption for most (if not all) intra-cellular cargos.

CONCLUSIONS AND OUTLOOK

This thesis has addressed a number of methodological and scientific questions pertaining to the study of endocytosis and trafficking of cargo in mammalian cells. The emphasis was on the development of computational techniques for image-based analysis, modeling, and simulation of biophysical processes that are linked to intra-cellular transport. Measurements that probe properties of the endocytic system in live cells are strongly constrained by physical processes, such as diffraction of light, which limits the resolution of optical microscopes. Model-based computational approaches are hence critical for the study of these phenomena.

Several key mechanisms and phenomena in endocytosis and trafficking depend on tight coupling between biophysical and biochemical processes. Active transport, for example, delivers internalized cargo to its destination, controls the spatial arrangement of organelles, affects the shape of intra-cellular objects, and regulates fusion and fission of membrane compartments, which is central for cargo sorting. In the systems studied in this thesis, relations to motor-dependent active transport are ubiquitous.

This thesis contributes novel data-analysis methods as well as statistical and physical models for studying endocytosis and intra-cellular trafficking. In particular, a novel algorithm was developed that decomposes heterogeneous trajectories into distinct segments containing only a single motion type. Such heterogeneous trajectories are frequently observed for objects moved by molecular motors, as demonstrated for virus motion inside cells and on their plasma membrane. The intra-cellular localization, shape, and dynamic shape changes of endosomes were studied using a novel image analysis technique that combines image deconvolution with active contour segmentation. The possibility of a coupling between virus infection and biophysical and biochemical causes of the shapes of endosomes was investigated using statistical tests on endosome shape feature distributions. Shape changes, as frequently caused by motor action, could be resolved. A statistical framework for studying spatial correlations between intra-cellular objects was introduced, leading to an extension of co-localization analysis to interaction analysis. Applying novel interaction measures and statistical tests allowed following virus trafficking through the endosomal network. A mechano-chemical model for cooperative, but unregulated cargo transport by several molecular motors was then presented. The model includes extensions to previous works that revealed non-trivial effects of motor–cargo coupling, as demonstrated in a large simulation study.

The key methodological and biological contributions of this thesis, limitations of the presented methods, and possibilities for future work are discussed in the following sections.

5.1 TRAJECTORY SEGMENTATION

In Chapter 1, a novel trajectory segmentation method was introduced and applied to trajectories of Adenovirus particles moving inside live cells and on the plasma membrane.

Adenovirus motion is highly heterogeneous with frequently alternating phases of directed motion, drifts, confinement, and diffusive motion. The duration of the different motion types could be as short as a few tenths of a second, which translated into some 10 consecutive point detections. The problem of detecting motion patterns was rephrased as a classifica-

tion problem on application-specific sets of motion features. Features were computed in overlapping windows over the trajectory. The representation of window contents by feature vectors is essential in overcoming the poor classification performance observed on raw point positions. The method's ability to reliably detect short segments of directed motion, enabling analysis of active cargo transport on the important one-second time scale, was demonstrated on synthetic data. The sensitivity and specificity of detecting different motion types were assessed on synthetic trajectories and manually segmented trajectories of Adenovirus motion and were found to be >90% for all motion types considered. In trajectories containing purely random motion, only negligible, insignificant amounts of the motion types were found.

The performance of trajectory segmentation was largely owed to the implicit definition of the criteria used to identify the different motion types. Support vector machines were trained with manually segmented trajectories. The training sets were iteratively augmented with additional samples where classification previously failed. Supervised classification circumvents the tedious process of manually defining classification criteria for motion types. The latter approach has been formerly applied to trajectories of secretory vesicle dynamics [65]. Furthermore, the present trajectory segmentation algorithm is highly flexible with respect to the detection of different motion types. No prior knowledge about the motion types is included in its design. Adaptation to specific motion types is achieved through the design of features and user-provided collections of training data.

Once trained, the trajectory segmentation algorithm can automatically segment large numbers of trajectories, as frequently produced in cell biological assays. The possibility to reliably and automatically segment large numbers of trajectories enabled frequency analyses for motion types in Adenovirus trajectories. A perturbation study unambiguously identified the dependence of directed and fast-drifting motion on microtubules.

Using the present trajectory segmentation algorithm, Gazzola and co-workers were able to extract velocity and run-length distributions subsequently used them to infer the parameters of an active transport model for Adenovirus particles [44].

In the presented benchmarks and applications, trajectory segments were reliably detected. Correct localization of transitions between different motion types, however, was not always guaranteed. The use of carefully trained supervised classifiers to transform a collection of features into a single vote greatly increases detection power, especially for short segments. Nevertheless, moving-window feature extraction has a strong averaging effect, which inevitably blurs transitions. Shorter feature-extraction windows in conjunction with advanced signal-processing methods for classifier output filtering could possibly increase transition localization accuracy. Therefore, a solid theoretical foundation for designing trajectory and classifier output filter functions should be developed on the basis of statistical decision theory.

5.2 DECONVOLVING ACTIVE CONTOUR SEGMENTATION OF INTRA-CELLULAR ORGANELLES

In Chapter 2, a novel image analysis method for segmenting intra-cellular objects from live cell fluorescence microscopy images was presented. The method was benchmarked on synthetic data, demonstrated on images of the Golgi apparatus and endosomes, and applied to quantifying the dynamic and static shapes of endosomes in virus-infected HER 911 cells.

Accurate segmentation of endosomes and similarly-sized intra-cellular organelles is challenging due to the limited optical resolution and signal-to-noise ratio available in live cell optical microscopy techniques. In the present work, this problem was addressed by incorporating the concept of deconvolution into image segmentation. Dimensionality reduction and regularization of the deconvolution problem was achieved by representing the solution as explicit active contours enclosing constant-intensity regions. As shown in tests on synthetic data, this allowed efficiently reconstructing details of outlines on the length scale of the width of the point spread function. The localization accuracy of the reconstructed outlines was on the order of tenths of pixels, even at modest signal-to-noise ratios. For signal-to-noise ratios larger than 10, the error of the estimated total intensity was below 5%.

Compared to classical pixel-based image segmentation techniques, deconvolving active contours reconstruct the true physical boundaries of the imaged objects in a less biased and more accurate way. As shown in applications considering endosome shapes in virus-infected cells and dynamic shape changes of during fusion of endosomes, this does not only improve the results quantitatively, but also reveals previously undetectable qualitative effects. Furthermore, the reconstructed outlines can be considered highly correlated with true object outlines, provided the model assumptions are valid. This high degree of correlation is of particular importance if the parameters of models of physical processes, such as the deformation of endosomes due to the collective action of motor proteins, have to be inferred from image data. In a recent study [133], it was concluded that the elongation of endosomes along the direction of microtubules is a result of a tug of war of dissimilar motors. Inference of, for example, the number of motors pulling an endosome requires precise quantification of the endosome deformation, as enabled by the image segmentation approach presented here.

In order to minimize the number of parameters to be estimated from the images, the fluorophore density within each imaged object was modeled with as constant. Different objects can have different intensities and intensities can change over time. An acceptable bias of outline estimates can only be expected if this model assumption is – at least approximately – valid. In the Golgi and endosome applications shown, the model was adequate, judging from the small residual error of fitted images. Allowing for more general fluorophore density models could potentially further the practical utility of the method. Even in the present simple form, however, inference of the fluorophore density model parameters is the dominant source of uncertainty. It is thus doubtful whether more complex models can be robustly inferred, without including prior knowledge about the true fluorophore density. A further limitation of the present method concerns the topology of the collection of reconstructed outlines, which is defined by the segmentation used for initializing the active contours. This segmentation can be ambiguous if objects are close to each other. An implicit representation of the outlines as level sets could circumvent this limitation [129, 90, 118, 93]. Nevertheless, the additional flexibility probably has to be constrained by additional shape or topology priors in order to avoid over-fitting.

5.3 INTERACTION ANALYSIS FOR INTRA-CELLULAR OBJECTS

A statistical inference framework for robustly estimating interaction parameters from experimentally observed spatial distributions of intra-cellular objects was introduced in Chapter 3.

For the first time, a connection between co-localization and spatial interaction could be established. This was achieved by phrasing the object-based interaction analysis problem in a spatial statistics framework that is based on nearest-neighbor distance distributions. This provided generic procedures for inferring interaction strengths and quantifying their statistical significance. Standard object-based co-localization analysis was included as a special case, making explicit the connections between spatial interaction and co-localization. Novel interaction measures were derived from a statistical model of nearest-neighbor distributions. These measures and the related statistical tests allowed following the trafficking of different Adenovirus strains through the endosomal network, in particular Rab5-positive early endosomes.

In the presented framework, two key quantities emerge: (i) the state density $q(d)$, which is the distribution of nearest-neighbor distances expected under the null hypothesis of no interaction, and (ii) the interaction potential $\phi(d)$, which defines the strength and distance dependence of the interaction. It was shown that classical co-localization analysis amounts to estimating the parameters of a step potential for the interaction. This requires a notion of “inside” and “outside”, either naturally defined by the physical extent of the objects or imposed through the step function’s distance threshold. For point-like objects, or weak correlations between object positions, the choice of distance threshold is, however, arbitrary.

These limitations were relaxed by allowing more general shapes of the interaction potential, which naturally extended co-localization analysis to (spatial) interaction analysis, without requiring any additional assumptions. The additional flexibility allows capturing information about a wider range of sub-cellular interactions. This was demonstrated by statistical power analysis of the classical and generalized measures. It was found that the probability of detecting an actual interaction strongly depends on the

5.3. INTERACTION ANALYSIS FOR INTRA-CELLULAR OBJECTS

cellular context, and that longer-ranging interactions are harder to detect. Test statistics that include prior knowledge about the shape of the true interaction potential can thus greatly reduce the number of samples required to achieve a certain target power. Physico-chemical models might provide such prior knowledge. Alternatively, a non-parametric phenomenological potential can be estimated from the data, as it was demonstrated here for the virus–endosome interaction.

The presented approach to interaction analysis enables applying a wide range of established statistical tools for analyzing experimental data, from parameter identification to model selection. This workflow was illustrated by studying the spatial patterns of endosomes and viruses infecting live human cells.

The presented framework is limited by the same assumptions that also underlie classical co-localization analysis: (i) spatial homogeneity and (ii) isotropy of the interaction within the observation window, and (iii) exclusively nearest-neighbor interactions between objects of different classes. Assumption (i) is, e.g., violated if large areas of the images analyzed do not contain any objects. In this case, estimation of $q(d)$ is not robust. Partitioning a cell’s interior into smaller sub-domains and estimating a hyper-model for the interactions within the individual sub-domains could potentially increase the robustness of parameter estimation in such cases and enable resolving finer spatial variations in the interaction process.

General Gibbs processes could be used to analyze many-body interactions. Such processes, however, are theoretically and numerically challenging.

The interpretation of fitted potentials is limited to their relative strengths. In the absence of a mechanistic or physical model of the process that has created the observed spatial pattern, biophysical interpretation of the identified parameter values is difficult or misleading. This is because the fitted interaction potentials reflect the collection of all intra-cellular phenomena that led to the observed point pattern. For the virus–endosome interaction considered here, these phenomena include passive (diffusion) and active (motor-dependent) transport of internalized viruses, structuring of the cytoplasm by collective dynamics of organelles and filaments [35, 75, 34, 33], fusion and fission of organelles during sorting processes [120, 33, 99], signal-induced penetration of organelle membranes by virus particles [114], etc.

Interestingly, however, a relation between the steady-state distribution of a diffusion process with deterministic drift and the distribution of the Gibbs process (Equation 3.6) exists: If the deterministic force causing the drift is given by $-\partial\phi/\partial d$, then the two distributions become identical (in appropriate units). This fact points a possibility of connecting fitted interaction potentials with biophysical models of transport processes.

5.4 ACTIVE TRANSPORT MODEL

A novel mechano-chemical model for cooperative – but uncoordinated – intra-cellular cargo transport by several molecular motors was presented in Chapter 4. The model was studied extensively using numerical simulations.

Molecular motors play important roles in many cellular functions, including endocytosis (discussed in Chapter 2) and intra-cellular trafficking of cargo (Chapter 1). Motors like, such as kinesin can sustain forces on the order of a few pico-Newton (10^{-12} N). Forces of this magnitude are sufficient to move small cargo, such as vesicles or virus particles, with velocities of some hundred nanometers per second through the highly viscous cytoplasm of cells. Furthermore, the lipid membranes of organelles may be significantly deformed under the action of motor forces, as, for example, is the case in membrane tubule extraction [77, 84] or fusion of endosomes. Pico-Newton forces also have a significant effect on the dissociation kinetics of motor–filament bonds. The coupling between deformation, drag, and the chemical reactions involved in the stochastic motor stepping renders the collective dynamics of motor–cargo complexes non-trivial; even in the absence of regulatory mechanisms.

In the presented model, the forces and their effect on reactions, such as stepping and unbinding of single motors were explicitly included. Due to the continuous motion of the cargo reaction rates continuously changed in-between reactions. Such hybrid stochastic-deterministic models can *exactly* and *efficiently* be simulated using a variant of the stochastic simulation algorithm. 1000 seconds of cargo motion, which required some 10^5 reactions, could be simulated in less than 1 second of CPU time using a sequential code running on a 2.66 GHz intel processor.

As demonstrated in a large-scale parameter study (based on >1000 trajectories lasting $2 \cdot 10^5$ seconds each), the drag coefficient γ of the cargo, which depends on cargo size, and the elasticity κ of the motor–cargo link, which mainly depends on cargo rigidity, had significant qualitative and quantitative influence on the transport properties. The sensitivity of the processivity of uni-directional transport with respect to the number of motors, binding rates, and motor forces has previously been demonstrated [108]. The present model showed that the elasticity of the motor–cargo link is of equal importance. Furthermore, the emergence of different mobility regimes, as characterized by the processivity of uni-directional motion, the duration of immobile deadlocks, and the number and values of distinct transport velocities, could be related to – and explained by – the magnitude of κ and γ .

Drag forces significantly reduce the velocity of larger intra-cellular cargo. In systems with several cooperating motors, the degree of multi modality of velocity distributions has been linked to cargo size [130]. Using the model presented in this thesis, this picture is extended by the observation that the elasticity of the motor–cargo link has a significant effect on the number of distinct velocities that the cargo can adopt. Interestingly, this number can both *increase* or *decrease* with increasing κ , depending on the values of the other model parameters. Several characteristics indeed depended on combinations of two or more parameters, and not simply single model parameters. This illustrates that interpreting observed cargo dynamics in the context of overly simple transport models may be misleading. Moreover, since the important model parameters κ and γ depend on physical properties of the cargo, results concerning specific systems (reviewed in References [54, 152, 45]), probably do not generalize across different cargos.

Future work is concerned with extending the present model to include cargo deformations and cargo–cargo interactions. The present model may be used as a basis for constructing models of intra-cellular transport dynamics on the scale of entire populations of organelles. Such models could fully resolve all internal degrees of freedom of individual motor–cargo complexes. The resulting simulations, however, would be computationally expensive. Alternatively, the present model can be used to estimate coarse-grained properties of single cargo objects and pairs of cargo

objects, such as effective unbinding rates, velocity distributions, effective stalling forces, or effective interaction potentials (similar to those discussed in Chapter 3) for cargo–cargo collisions. Using these properties, a coarse-grained population-level model could be constructed that enables more efficient simulations of collective cargo behavior. This collective behavior in turn forms the basis for the theory of spatial patterns of intra-cellular organelles.

5.5 SIGNIFICANCE OF THE PRESENT WORK

This thesis has presented work at the interface between biology and computational science: Biological questions are tackled by, e.g., analyzing complex data using computational tools, performing numerical simulations of accepted models, or creating new models from unstructured pieces of information.

In the process of modeling, false positive and false negative are a major concern. If one uses statistical tests with a prescribed significance level of, say, 5%, one out of 20 trials is expected to be a false positive. In practice this means that if in only few cases the behavior of the studied system is unexpected and failure to detect this is likely one has to expect that a large portion of all rejected hypotheses are actually not false. It is therefore important to work with high detection power. This can be achieved by collecting more data, better statistical tests, or reduction of systematic and unsystematic errors in the data¹. A large portion of the work presented in this thesis has pushed the state of the art in these directions by enabling automated (by using efficient software), accurate (by reducing [human] bias), and precise (by using all available information in the data) analysis of large sets of biological data pertaining to intra-cellular transport.

The general significance of this thesis is that the newly developed algorithms and models can help structure and systematize research in cell biology. Furthermore, the presented algorithms improve the analysis of frequently encountered data and may motivate novel experimental approaches.

¹Note that decreasing the significance level is usually not an option.

Trajectory segmentation: Vast numbers of trajectories of intra-cellular objects are nowadays routinely acquired. Their systematic analysis is hampered by factors such as alternating motion patterns, dilution of the data sets by spurious detections (no real objects) or objects not moving at all, or erroneous links introduced by the particle tracking algorithm. The trajectory segmentation algorithm presented here allows “cleaning” data sets from stretches of uninteresting motion or corrupt trajectories in an automated way. This is essential for accurate and precise quantification of physical parameters (or any other statistic such as MSD or MSS) of the studied transport processes, as the “uninteresting” trajectory parts bias statistics and increase their variance. The algorithm is easy to use and adapt, and it can be used in high-throughput screens, thanks to its simplicity and efficiency.

Reconstruction of organelle shapes: The image analysis algorithm presented in Chapter 2 combines ideas from deconvolution and image segmentation. So far, the necessity to account for the strong blurring introduced by fluorescence microscopy optics was largely ignored in image segmentation. The output of classical image segmentation techniques was frequently used as an estimate of the imaged object’s shape. For small intra-cellular structures, these estimates are strongly biased and imprecise. The novel algorithm reduces these deficiencies, which opens new possibilities for automated, high-throughput, and unbiased characterization of the morphology of intra-cellular objects. This may provide the high-quality data required to unravel biophysical causes of organelle shapes. Furthermore, the automatic detection of fusion events may be used to map the connectivity of trafficking networks from live cell imaging experiments. In addition to motivating such biological experiments, the algorithm points to possibilities for future research at the largely unexplored interface of image deconvolution and segmentation.

Image-based interaction analysis: The statistical framework introduced in Chapter 3 provides a solid theoretical foundation for the analysis of spatial patterns of intra-cellular objects. Such patterns are very common data that are frequently used to infer functional relations between intra-cellular objects. The spatial proximity required for direct molecular interactions, and the high degree of compartmentalization of eukaryotic cells, previously motivated simple overlap co-localization measures. Such

measures, however, are just one instance of the general class of measures derived here from a statistical model for the spatial patterns of objects. Within this model, more general correlations in the localization patterns of two classes of objects are assumed to be caused by an effective interaction between the objects. Most importantly, prior knowledge about the studied system can be formally described by a hypothesis for the shape of the effective interaction potential. This hypothesis can then directly be tested against the observed localization pattern. The presented framework provides statistical tools such as parameter estimation, model selection, non-parametric hypothesis tests, etc. Moreover, it can assist in the design of experiments by, e.g., a-priori characterization of the detection power of an experimental test of a hypothesis.

Model for active intra-cellular transport: Intra-cellular transport is a complex process that is still not fully understood, although many elements of the transport machinery have been identified and characterized. This is partly due to the fact that the bigger picture is hard to see from the individual pieces of knowledge. The model presented in Chapter 4 serves as a container to organize knowledge about the different aspects and elements of intra-cellular active transport. Using computer simulations of this model or variants thereof, one can test new hypotheses about the working mechanisms of intra-cellular transport. Moreover, the model allows measuring non-observable variables and provides control over experimentally un-controllable parameters. Prior to experimental assays, the model's output can be explored in order to identify the most informative experimental conditions and testable hypotheses. Several important parameters of the model, such as single motor properties, can be estimated from optical trap experiments. Links to particle tracking experiments can be established through the physical properties of cargo motion, such as the average velocity or the length of uninterrupted runs along filaments. By reproducing experimental distributions of these properties, unknown parameters of the model can possibly be estimated. A potentially interesting readout that could be reproduced by a simple extension of the present model is the shape of deformable cargo as caused by motor forces. Using the approaches presented in Chapters 1 and 2, such shapes could be reconstructed and tracked from fluorescence microscopy images, and the resulting trajectories could be segmented in order to focus the analysis on, e.g., filament-dependent motion types. Such data could be used to further

test the validity of different model assumptions. Finally, the model can be extended to realistically resolve cargo–cargo interactions in order to build a model of entire populations of cargo. This population-level model would generate spatial patterns of organelles and could serve as a starting point for a biophysical interpretation of the effective interaction potentials estimated in Chapter 3.

BIBLIOGRAPHY

- [1] S. S. Andrews and D. Bray. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.*, 1:137–151, 2004.
- [2] E. Anlauf and A. Derouiche. A practical calibration procedure for fluorescence colocalization at the single organelle level. *J. Microsc.*, 233:225–233, 2009.
- [3] I. Aprodu, M. Soncini, and A. Redaelli. Mechanical characterization of motor proteins: A molecular dynamics approach. *Macromol. Theor. Simul.*, 17(7–8):376–384, 2008.
- [4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans. Signal Process.*, 50(2):174–188, 2002.
- [5] R. Assunção. Score test for pairwise interaction parameters of gibbs point processes. *Brazilian Journal of Probability and Statistics*, 17:169–178, 2003.
- [6] A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics*, 42(3):283–315, 2000.

BIBLIOGRAPHY

- [7] M. Badoual, F. Jülicher, and J. Prost. Bidirectional cooperative motion of molecular motors. *Proc. Natl. Acad. Sci. USA*, 99(10):6696–6701, 2002.
- [8] W. Baumeister. From proteomic inventory to architecture. *FEBS Letters*, 579(4):933–937, Feb. 2005.
- [9] J. Beeg, S. Klumpp, R. Dimova, R. S. Gracia, E. Unger, and R. Lipowsky. Transport of beads by several kinesin motors. *Biophys. J.*, 94(2):532–541, Jan. 2008.
- [10] M. Berman and T. R. Turner. Approximating point process likelihoods with GLIM. *Applied Statistics-Journal of the Royal Statistical Society Series C*, 41(1):31–38, 1992.
- [11] M. Berry and Z. Lewis. On the Weierstrass-Mandelbrot fractal function. *P. Roy. Soc. Lond. A Mat.*, 370(1743):459–484, 1980.
- [12] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1 edition, 1996.
- [13] S. Bolte and F. P. Cordelieres. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.*, 224:213–232, 2006.
- [14] A. Borodin and P. Salminen. *Handbook of Brownian Motion – Facts and Formulae*. Birkhäuser, 2nd edition, 2002.
- [15] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 1st edition, 2004.
- [16] C. J. Burckhardt and U. F. Greber. Virus movements on the plasma membrane support infection and transmission between cells. *PLoS Pathog.*, 5(11), 2009.
- [17] K. P. Burnham and D. Anderson. *Model Selection and Multi-Model Inference*. Springer, 2nd edition, 2002.
- [18] C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky. Mechanical processes in biochemistry. *Annu. Rev. Biochem.*, 73:705–748, 2004.

-
- [19] O. Campàs, Y. Kafri, K. B. Zeldovich, J. Casademunt, and J.-F. Joanny. Collective dynamics of interacting molecular motors. *Phys. Rev. Lett.*, 97(3):038101, 2006.
- [20] M. B. Cannell, A. McMorland, and C. Soeller. *Handbook of Biological Confocal Microscopy*, chapter 25: Image enhancement by deconvolution. Springer, New York, 2006.
- [21] N. Carter and R. Cross. Mechanics of the kinesin step. *Nature*, 435(7040):308–312, May 2005.
- [22] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Process.*, 10:266–277, 2001.
- [23] M. K. Cheezum, W. F. Walker, and W. H. Guilford. Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophys. J.*, 81:2378–2388, 2001.
- [24] D. Chowdhury, A. Schadschneider, and K. Nishinari. Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms. *Phys. Life Rev.*, 2(4):318–352, 2005.
- [25] L. D. Cohen. On active contour models and balloons. *Cvgip: Image Understanding*, 53(2):211–218, Mar. 1991.
- [26] C. Collinet, M. Stoeter, C. R. Bradshaw, N. Samusik, J. C. Rink, D. Kenski, B. Habermann, F. Buchholz, R. Henschel, M. S. Mueller, W. E. Nagel, E. Fava, Y. Kalaidzidis, and M. Zerial. Systems survey of endocytosis by multiparametric image analysis. *Nature*, 464(7286):243–249, 2010.
- [27] S. Costes, D. Daelemans, E. Cho, Z. Dobbin, G. Pavlakis, and S. Lockett. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys. J.*, 86(6):3993–4003, June 2004.
- [28] M. Cotten and J. Weber. The adenovirus protease is required for virus entry into host-cells. *Virology*, 213:494–502, 1995.
- [29] L. Coulot, H. Kirschner, A. Chebira, J. M. F. Moura, J. Kovačević, E. G. Osuna, and R. F. Murphy. Topology preserving STACS segmentation of protein subcellular location images. *Proc. IEEE Int. Symp. Biomed. Imaging*, pages 566–569, 2006.

BIBLIOGRAPHY

- [30] H. Delingette and J. Montagnat. Shape and topology constraints on parametric active contours. *Computer Vision and Image Understanding*, 83(2):140–171, Aug. 2001.
- [31] A. Desai and T. J. Mitchison. Microtubule polymerization dynamics. *Annu. Rev. Cell Dev. Biol.*, 13:83–117, 1997.
- [32] P. Diggle. *Statistical Analysis of Spatial Point Patterns*. A Hodder Arnold Publication, 2nd edition, 2003.
- [33] A.-T. Dinh, C. Pangarkar, and S. Mitragotri. Understand discrete nanoscale transport. *Chem. Eng. Prog.*, 104(6):62–68, 2008.
- [34] A.-T. Dinh, C. Pangarkar, T. Theofanous, and S. Mitragotri. Theory of spatial patterns of intracellular organelles. *Biophys. J.*, 90(10):L67–L69, 2006.
- [35] A. T. Dinh, T. Theofanous, and S. Mitragotri. A model for intracellular trafficking of adenoviral vectors. *Biophys. J.*, 89(3):1574–1588, 2005.
- [36] M. Dogterom, J. Kerssemakers, G. Romet-Lemonne, and M. Janson. Force generation by dynamic microtubules. *Curr. Opin. Cell Biol.*, 17:67–74, 2005.
- [37] R. O. Duda, H. P. E., and S. D. G. *Pattern Classification*. Wiley Interscience, 2 edition, 2000.
- [38] A. Dufour, V. Shinin, S. Tajbakhsh, N. Guillen-Aghion, J. C. Olivo-Marin, and C. Zimmer. Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. *IEEE Trans. Image Process.*, 14:1396–1410, 2005.
- [39] N. El-Zehiry and A. Elmaghraby. An active surface model for volumetric image segmentation. *Proc. IEEE Int. Symp. Biomed. Imaging*, pages 1358–1361, June 2009.
- [40] R. Ferrari, A. J. Manfroi, and W. R. Young. Strongly and weakly self-similar diffusion. *Physica D*, 154:111–137, 2001.
- [41] A. P. French, S. Mills, R. Swarup, M. J. Bennett, and T. P. Pridmore. Colocalization of fluorescent markers in confocal microscope images of plant cells. *Nature Protocols*, 3(4):619–628, 2008.

-
- [42] L. Gao, R. Lipowsky, and J. Shillcock. Tension-induced vesicle fusion: pathways and pore dynamics. *Soft Matter*, 4(6):1208–1214, 2008.
- [43] M. Gastaldelli, N. Imelli, K. Boucke, B. Amstutz, O. Meier, and U. F. Greber. Infectious adenovirus type 2 transport through early but not late endosomes. *Traffic*, 9:2265–2278, 2008.
- [44] M. Gazzola, C. J. Burckhardt, B. Bayati, M. Engelke, U. F. Greber, and P. Koumoutsakos. A stochastic model for microtubule motors describes the in vivo cytoplasmic transport of human adenovirus. *PLoS Comp. Biol.*, 5(12), Dec. 2009.
- [45] A. Gennerich and R. D. Vale. Walking the walk: how kinesin and dynein coordinate their steps. *Curr. Opin. Cell Biol.*, 21(1):59–67, 2009.
- [46] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104:1876–1889, 2000.
- [47] S. F. Gibson and F. Lanni. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *J. Opt. Soc. Am. A*, 8(10):1601–1613, 1991.
- [48] B. N. G. Giepmans, S. R. Adams, M. H. Ellisman, and R. Y. Tsien. Review - the fluorescent toolbox for assessing protein location and function. *Science*, 312:217–224, 2006.
- [49] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.
- [50] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Ann. Rev. Phys. Chem.*, 58:35–55, 2007.
- [51] U. F. Greber and M. Way. A superhighway to virus infection. *Cell*, 124(4):741–754, 2006.
- [52] U. F. Greber, P. Webster, J. Weber, and A. Helenius. The role of the adenovirus protease in virus entry into cells. *EMBO J.*, 15:1766–1777, 1996.

BIBLIOGRAPHY

- [53] U. F. Greber, M. Willetts, P. Webster, and A. Helenius. Stepwise dismantling of adenovirus-2 during entry into cells. *Cell*, 75:477–486, 1993.
- [54] S. P. Gross. Hither and yon: a review of bi-directional microtubule-based transport. *Phys. Biol.*, 1(1-2):R1–R11, 2004.
- [55] Y. Guo and A. D. Linstedt. COPII-Golgi protein interactions regulate COPII coat assembly and Golgi size. *J. Cell Biol.*, 174:53–63, 2006.
- [56] N. Hansen. *The CMA Evolution Strategy: A Tutorial*, Apr 2007.
- [57] N. Hansen and A. Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In *Proceedings of the 1996 IEEE Conference on Evolutionary Computation (ICEC '96)*, pages 312–317, 1996.
- [58] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 1 edition, July 30 2003.
- [59] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 3rd edition, 2009.
- [60] M. Hayashi, A. Raimondi, E. O’Toole, S. Paradise, C. Collesi, O. Cremona, S. M. Ferguson, and P. De Camilli. Cell- and stimulus-dependent heterogeneity of synaptic vesicle endocytic recycling mechanisms revealed by studies of dynamin 1-null neurons. *Proc. Natl. Acad. Sci. USA*, 105:2175–2180, 2008.
- [61] E. Hecht. *Optics*. Addison, 4th edition, 2002.
- [62] J. Heikkinen and A. Penttinen. Bayesian smoothing in the estimation of the pair potential function of gibbs point processes. *Bernoulli*, 5:1119–1136, 1999.
- [63] T. J. Holmes, D. Biggs, and A. Abu-Tarif. *Handbook of Biological Confocal Microscopy*, chapter 24: Blind Deconvolution. Springer, New York, 2006.

-
- [64] G. Holzwarth, K. Bonin, and D. B. Hill. Forces required of kinesin during processive transport through cytoplasm. *Biophys. J.*, 82(4):1784–1790, 2002.
- [65] S. Huet, E. Karatekin, V. S. Tran, I. Fanget, S. Cribier, and J.-P. Henry. Analysis of transient behavior in complex trajectories: Application to secretory vesicle dynamics. *Biophys. J.*, 91:3542–3559, 2006.
- [66] N. Imelli, Z. Ruzsics, D. Puntener, M. Gastaldelli, and U. F. Greber. Genetic reconstitution of the human adenovirus type 2 temperature-sensitive 1 mutant defective in endosomal escape. *Virology Journal*, 6(174), 2009.
- [67] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser. Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods*, 5(8):695–702, 2008.
- [68] E. A. Jares-Erijman and T. M. Jovin. Imaging molecular interactions in living cells by FRET microscopy. *Curr. Opin. Chem. Biol.*, 10(5):409–416, 2006.
- [69] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.
- [70] S. Jeney, E. H. K. Stelzer, H. Grubmuller, and E. L. Florin. Mechanical properties of single motor molecules studied by three-dimensional thermal force probing in optical tweezers. *Chemphyschem*, 5(8):1150–1158, 2004.
- [71] J. R. Kardon and R. D. Vale. Regulators of the cytoplasmic dynein motor. *Nat. Rev. Mol. Cell Bio.*, 10(12):854–865, 2009.
- [72] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. Computer Vision*, pages 321–331, 1988.
- [73] D. Keller and C. Bustamante. The mechanochemistry of molecular motors. *Biophys. J.*, 78(2):541–556, 2000.
- [74] S. Klumpp and R. Lipowsky. Cooperative cargo transport by several molecular motors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(48):17284–17289, Nov. 2005.

BIBLIOGRAPHY

- [75] S. Klumpp, T. M. Nieuwenhuizen, and R. Lipowsky. Self-organized density patterns of molecular motors in arrays of cytoskeletal filaments. *Biophys. J.*, 88(5):3118–3132, 2005.
- [76] A. B. Kolomeisky and M. E. Fisher. Molecular motors: A theorist’s perspective. *Ann. Rev. Phys. Chem.*, 58:675–695, 2007.
- [77] G. Koster, M. VanDuijn, B. Hofs, and M. Dogterom. Membrane tube formation from giant vesicles by dynamic association of motor proteins. *Proc. Natl. Acad. Sci. USA*, 100(26):15583–15588, 2003.
- [78] M. Kozubek and P. Matula. An efficient algorithm for measurement and correction of chromatic aberrations in fluorescence microscopy. *J. Microsc.*, 200:206–217, 2000.
- [79] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.
- [80] A. Kunwar and A. Mogilner. Robust transport by multiple motors with nonlinear force-velocity relations and stochastic load sharing. *Phys. Biol.*, 7(1), 2010.
- [81] A. Kunwar, M. Vershinin, J. Xu, and S. P. Gross. Stepping, strain gating, and an unexpected force-velocity curve for multiple-motor-based transport. *Curr. Biol.*, 18(16):1173–1183, 2008.
- [82] A. Kusumi, C. Nakada, K. Ritchie, K. Murase, K. Suzuki, H. Murakoshi, R. S. Kasai, J. Kondo, and T. Fujiwara. Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: High-speed single-molecule tracking of membrane molecules. *Annu. Rev. Bioph. Biom. Struct.*, 34:351–378, 2005.
- [83] E. Lachmanovich, D. Shvartsman, Y. Malka, C. Botvin, Y. Henis, and A. Weiss. Co-localization analysis of complex formation among membrane proteins by computerized fluorescence microscopy: application to immunofluorescence co-patching studies. *J. Microsc.*, 212(2):122–131, 2003.
- [84] C. Leduc, O. Campas, K. B. Zeldovich, A. Roux, P. Jolimaitre, L. Bourel-Bonnet, B. Goud, J. F. Joanny, P. Bassereau, and J. Prost.

-
- Cooperative extraction of membrane nanotubes by molecular motors. *Proc. Natl. Acad. Sci. USA*, 101(49):17096–17101, 2004.
- [85] M. J. Lehmann, N. M. Sherer, C. B. Marks, M. Pypaert, and W. Mothes. Actin- and myosin-driven movement of viruses along filopodia precedes their entry into cells. *J. Cell Biol.*, 170(2):317–25, 2005.
- [86] V. Levi, A. S. Serpinskaya, E. Gratton, and V. Gelfand. Organelle transport along microtubules in xenopus melanophores: Evidence for cooperation between multiple motors. *Biophys. J.*, 90(1):318–327, 2006.
- [87] H. Li, T. Shen, M. B. Smith, I. Fujiwara, D. Vavylonis, and X. Huang. Automated actin filament segmentation, tracking and tip elongation measurements based on open active contours. *Proc. IEEE Int. Symp. Biomed. Imaging*, pages 1302–1305, June 2009.
- [88] D. C. Liebler. *Introduction to Proteomics: Tools for the New Biology*. Humana Press, 1st edition, 2002.
- [89] E. H. Linfoot and E. Wolf. Phase distribution near focus in an aberration-free diffraction image. *Proc. Phys. Soc. B*, 69(8):823–832, 1956.
- [90] D. Lingrand and J. Montagnat. Levelset and B-spline deformable model techniques for image segmentation: A pragmatic comparative study. In *Image Analysis*, volume 3540 of *Lecture Notes in Computer Science*, pages 25–34, 2005.
- [91] V. A. Lizunov, H. Matsumoto, J. Zimmerberg, S. W. Cushman, and V. A. Frolov. Insulin stimulates the halting, tethering, and fusion of mobile glut4 vesicles in rat adipose cells. *J. Cell Biol.*, 169(3):481–489, 2005.
- [92] V. Ljosa and A. E. Carpenter. Introduction to the quantitative analysis of two-dimensional fluorescence microscopy images for cell-based screening. *PLoS Comp. Biol.*, 5(12), Dec. 2009.
- [93] M. Machacek and G. Danuser. Morphodynamic profiling of protrusion phenotypes. *Biophys. J.*, 90:1439–1452, 2006.

BIBLIOGRAPHY

- [94] R. Mallik and S. P. Gross. Molecular motors: Strategies to get along. *Current Biology*, 14(22):971–982, Nov. 2004.
- [95] R. Mallik and S. P. Gross. Molecular motors as cargo transporters in the cell - the good, the bad and the ugly. *Physica A*, 372(1):65–69, 2006.
- [96] E. Manders, F. J. Verbeek, and J. A. Aten. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.*, 169(3), 1993.
- [97] M. Marsh and R. Bron. SFV infection in CHO cells: Cell-type specific restrictions to productive virus entry at the cell surface. *J. Cell Sci.*, 110:95–103, 1997.
- [98] S. Mayor and R. E. Pagano. Pathways of clathrin-independent endocytosis. *Nature Reviews Molecular Cell Biology*, 8(8):603–612, 2007.
- [99] L. S. Mayorga and E. M. Campoy. Modeling fusion/fission-dependent intracellular transport of fluid phase markers. *Traffic*, 2010 (accepted article).
- [100] J. G. McNally, T. Karpova, J. Cooper, and J. A. Conchello. Three-dimensional imaging by deconvolution microscopy. *Methods Enzymol.*, 19(3):373–385, 1999.
- [101] N. A. Medeiros, D. T. Burnette, and P. Forscher. Myosin II functions in actin-bundle turnover in neuronal growth cones. *Nat. Cell Biol.*, 8(3):215–26, 2006.
- [102] O. Meier and U. F. Greber. Adenovirus endocytosis. *J. Gene Med.*, 5(6):451–462, 2004.
- [103] E. Meijering, I. Smal, O. Dzyubachyk, and J.-C. Olivo-Marin. *Microscope Image Processing*, chapter 15: Time Lapse Imaging. Academic Press, 2008.
- [104] I. Mellman and G. Warren. The road taken: Past and future foundations of membrane traffic. *Cell*, 100:99–112, 2000.
- [105] M. Miaczynska, L. Pelkmans, and M. Zerial. Not just a sink: endosomes in control of signal transduction. *Curr. Opin. Cell Biol.*, 16:400–406, 2004.

-
- [106] Z. Michalewicz and D. B. Fogel. *How to Solve It: Modern Heuristics*. Springer, 3rd edition, 2002.
- [107] J. Møller and R. Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2004.
- [108] M. J. I. Müller, S. Klumpp, and R. Lipowsky. Tug-of-war as a co-operative mechanism for bidirectional cargo transport by molecular motors. *Proceedings of the National Academy of Sciences of the United States of America*, 105(12):4609–4614, Mar. 2008.
- [109] B. Neumann, M. Held, U. Liebel, H. Erfle, P. Rogers, R. Pepperkok, and J. Ellenberg. High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat. Methods*, 3(5):385–390, 2006.
- [110] S. Nickell, C. Kofler, A. P. Leis, and W. Baumeister. A visual approach to proteomics. *Nat. Rev. Mol. Cell Bio.*, 7:225–230, 2006.
- [111] J. Nocedal and W. S. J. *Numerical Optimization*. Springer, 2nd edition, 2006.
- [112] N. Numata, T. Kon, T. Shima, K. Imamuta, T. Mogami, R. Ohkura, K. Sutoh, and K. Sutoh. Molecular mechanism of force generation by dynein, a molecular motor belonging to the aaa plus family. *Biochemical Society Transactions*, 36:131–135, Feb. 2008.
- [113] L. Pelkmans. Viruses as probes for systems analysis of cellular signalling, cytoskeleton reorganization and endocytosis. *Curr. Opin. Microbiol.*, 8(3):331–337, 2005.
- [114] L. Pelkmans and A. Helenius. Insider information: what viruses tell us about endocytosis. *Curr. Opin. Cell Biol.*, 15:414–422, 2003.
- [115] H. Peng. Bioimage informatics: a new area of engineering biology. *Bioinformatics*, 24:1827–1836, 2008. Bioinformatics.
- [116] K. D. G. Pflieger and K. A. Eidne. Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat. Methods*, 3(3):165–174, 2006.
- [117] R. C. Piper and D. J. Katzmann. Biogenesis and function of multi-vesicular bodies. *Annu. Rev. Cell Dev. Biol.*, 23:519–547, 2007.

BIBLIOGRAPHY

- [118] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi. Particle filtering for geometric active contours with application to tracking moving and deforming objects. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2–9, 2005.
- [119] S. Rice, A. W. Lin, D. Safer, C. L. Hart, N. Naber, B. O. Carragher, S. M. Cain, E. Pechatnikova, E. M. Wilson-Kubalek, M. Whittaker, E. Pate, R. Cooke, E. W. Taylor, R. A. Milligan, and R. D. Vale. A structural change in the kinesin motor protein that drives motility. *Nature*, 402(6763):778–784, 1999.
- [120] J. Rink, E. Ghigo, Y. Kalaidzidis, and M. Zerial. Rab conversion as a mechanism of progression from early to late endosomes. *Cell*, 122:735–749, 2005.
- [121] R. L. Roberts, M. A. Barbieri, K. M. Pryse, M. Chua, and P. D. Stahl. Endosome fusion in living cells overexpressing GFP-rab5. *J. Cell Sci.*, 112:3667–3675, 1999.
- [122] P. Sarder and A. Nehorai. Deconvolution methods for 3-D fluorescence microscopy images. *IEEE Signal Process. Mag.*, 23:32–45, 2006.
- [123] M. Saxton. Anomalous subdiffusion in fluorescence photobleaching recovery: A monte carlo study. *Biophys. J.*, 81:2226–2240, 2001.
- [124] M. J. Saxton. *Fundamental Concepts in Biophysics: Volume 1 (Handbook of Modern Biophysics)*, chapter 6: Single-Particle Tracking. Humana Press, 2009.
- [125] I. F. Sbalzarini and P. Koumoutsakos. Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.*, 151(2):182–195, 2005.
- [126] I. F. Sbalzarini, A. Mezzacasa, A. Helenius, and P. Koumoutsakos. Effects of organelle shape on fluorescence recovery after photobleaching. *Biophys. J.*, 89(3):1482–1492, 2005.
- [127] B. Schölkopf and A. J. Smola. *Learning with Kernels. Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, Cambridge, Massachusetts, 2002.

-
- [128] G. Scita and P. P. Di Fiore. The endocytic matrix. *Nature*, 463(7280):464–473, 2010.
- [129] J. A. Sethian. Evolution, implementation, and application of level set and fast marching methods for advancing fronts. *J. Comput. Phys.*, 169:503–555, 2001.
- [130] Y. Shtridelman, T. Cahyuti, B. Townsend, D. DeWitt, and J. C. Macosko. Force-velocity curves of motor proteins cooperating in vivo. *Cell Biochem. Biophys.*, 52(1):19–29, 2008.
- [131] J. B. Sibarita. Deconvolution microscopy. *Adv. Biochem. Eng. Biot.*, 95:201–243, 2005.
- [132] J. S. Simonoff. *Smoothing Method in Statistics*. Springer, New York, 1st edition, 1998.
- [133] V. Soppina, A. K. Rai, A. J. Ramaiya, P. Barak, and R. Mallik. Tug-of-war between dissimilar teams of microtubule motors regulates transport and fission of endosomes. *Proc. Natl. Acad. Sci. USA*, 106(46):19381–19386, 2009.
- [134] J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons, 1st edition, 2003.
- [135] G. Srinivasa, M. C. Fickus, M. N. Gonzalez-Rivero, S. Y. Hsieh, Y. Guo, A. D. Linstedt, and J. Kovačević. Active mask segmentation for the cell-volume computation and Golgi-body segmentation of HeLa cell images. *Proc. IEEE Int. Symp. Biomed. Imaging*, pages 348–351, May 2008.
- [136] G. Srinivasa, M. C. Fickus, Y. Guo, A. D. Linstedt, and J. Kovačević. Active mask segmentation of fluorescence microscope images. *IEEE Trans. Image Process.*, 18(8):1817–1829, 2009.
- [137] H. Stenmark. Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Bio.*, 10(8):513–525, 2009.
- [138] D. J. Stephens and V. J. Allan. Light microscopy techniques for live cell imaging. *Science*, 300:82–86, 2003.

BIBLIOGRAPHY

- [139] S. R. Sternberg. Grayscale morphology. *Comput. Vis. Graph.*, 35:333–355, 1986.
- [140] D. Stoyan and A. Penttinen. Recent applications of point process methods in forestry statistics. *Statistical Science*, 15:61–78, 2000.
- [141] G. J. Streekstra and J. van Pelt. Analysis of tubular structures in three-dimensional confocal images. *Network: Comput. Neural Syst.*, 13:381–395, 2002.
- [142] M. Suomalainen, M. Y. Nakano, S. Keller, K. Boucke, R. P. Stidwill, and U. F. Greber. Microtubule-dependent plus- and minus end-directed motilities are competing processes for nuclear targeting of adenovirus. *J. Cell Biol.*, 144(4):657–672, 1999.
- [143] K. Svoboda and S. M. Block. Force and velocity measured for single kinesin molecules. *Cell*, 77(5):773–784, June 1994.
- [144] D. Toomre and D. J. Manstein. Lighting up the cell surface with evanescent wave microscopy. *Trends Cell Biol.*, 11(7):298–303, 2001.
- [145] A. M. Turing. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. London*, B237:37–72, 1952.
- [146] R. D. Vale. The molecular motor toolbox for intracellular transport. *Cell*, 112(4):467–480, 2003.
- [147] G. M. P. van Kempen, L. J. van Vliet, P. J. Verveer, and H. T. M. van der Voort. A quantitative comparison of image restoration methods for confocal microscopy. *J. Microsc.*, 185:354–365, 1997.
- [148] A. Vilfan, E. Frey, and F. Schwabl. Elastically coupled molecular motors. *Eur. Phys. J. B*, 3:535–546, 1998.
- [149] L. Vincent and P. Soille. Watersheds in digital spaces - an efficient algorithm based on immersion simulations. *IEEE T. Pattern. Anal.*, 13:583–598, 1991.
- [150] A. Vonderheit and A. Helenius. Rab7 associates with early endosomes to mediate sorting and transport of semliki forest virus to late endosomes. *PLoS Biol.*, 3:1225–1238, 2005.

-
- [151] M. Weiss, H. Hashimoto, and T. Nilsson. Anomalous protein diffusion in living cells as seen by fluorescence correlation spectroscopy. *Biophys. J.*, 84:4043–4052, 2003.
- [152] M. A. Welte. Bidirectional transport along microtubules. *Curr. Biol.*, 14(13):525–537, Jul 2004.
- [153] S. Wieser, M. Axmann, and G. J. Schütz. Versatile analysis of single-molecule tracking data by comprehensive testing against monte carlo simulations. *Biophys. J.*, 95(12):5988–6001, 2008.
- [154] B. Zhang, N. Chenouard, J.-C. Olivio-Marin, and V. Meas-Yedid. Statistical colocalization in biological imaging with false discovery control. *IEEE International Symposium on Biomedical Imaging, ISBI 2008, Paris, France*, pages 1327–1330, 2008.
- [155] V. Zinchuk, O. Zinchuk, and T. Okada. Quantitative colocalization analysis of multicolor confocal immunofluorescence microscopy images: Pushing pixels to explore biological phenomena. *Acta Histochemica Et Cytochemica*, 40(4):101–111, 2007.

LIST OF OWN PUBLICATIONS

Parts of this thesis have been published in the following papers:

Refereed journal papers

1. J. A. Helmuth, C. J. Burckhardt, P. Koumoutsakos, U. F. Greber, and I. F. Sbalzarini. A novel supervised trajectory segmentation algorithm identifies distinct types of human adenovirus motion in host cells. *J. Struct. Biol.* 159(3):347–358, 2007.
2. J. A. Helmuth, C. J. Burckhardt, U. F. Greber, and I. F. Sbalzarini. Shape reconstruction of subcellular structures from live cell fluorescence microscopy images. *J. Struct. Biol.* 167(1):1–10, 2009.
3. J. A. Helmuth, G. Paul, and I. F. Sbalzarini. Beyond co-localization: inferring spatial interactions between sub-cellular structures from microscopy images. *BMC Bioinf.* 11:372, 2010.

Refereed conference papers

1. J. A. Helmuth and I. F. Sbalzarini. Deconvolving active contours for fluorescence microscopy images. In *Proc. Intl. Symp. Visual Computing (ISVC)*, volume 5875 of Lecture Notes in Computer Science, pages 544–553, Las Vegas, USA, November 2009. Springer.

CURRICULUM VITAE

JO ARNE HELMUTH

ADDRESS AND PERSONAL DETAILS

Address: Meinrad Lienert-Strasse 15
CH-8003 Zurich
Switzerland
Phone: +41-44-4504828
E-mail: jo.helmuth@inf.ethz.ch
Date of Birth: July 20th, 1979, Hannover, Germany
Nationality: German

EDUCATION

1986 – 1992 Primary school in Hannover, Germany
1992 – 1999 High School in Hannover, Germany (qualification for university entrance)
2000 – 2002 Basic study period of Mechanical Engineering at ETH Zurich
2002 – 2005 Advanced study period of Mechanical Engineering at ETH Zurich; Specialization on “Computational Science” and “Nanotechnology”
2006 – 2010 PhD studies; Prof. Ivo Sbalzarini, Institute of Theoretical Computer Science, ETH Zurich

PRACTICAL EXPERIENCE

08/99 – 06/00 Civilian service; “Deutsches Rotes Kreuz”, Hannover, Germany
02/01 – 03/01 Internship; “Continental AG”, Hannover, Germany
10/02 – 07/03 Part time teaching assistant; Institute of Mechanics, ETH Zurich

11/03 – 02/04	Internship; “ABB Schweiz AG”, Baden, Switzerland
02/05 – 12/05	Part time research assistant; Institute of Zoology, University of Zurich
04/05 – 07/05	Diploma project at IBM Research GmbH, Zurich Research Laboratory, Switzerland
01/06 – 06/06	Research assistant; Institute of Computational Science, ETH Zurich

SCIENTIFIC PUBLICATIONS

J. A. Helmuth, H. Schmid, R. Stutz, A. Stemmer, and H. Wolf. High-speed microcontact printing. *J. Am. Chem. Soc.* 2006; 128: 9296–9297

J. A. Helmuth, C. J. Burckhardt, P. Koumoutsakos, U. F. Greber, and I. F. Sbalzarini. A novel supervised trajectory segmentation algorithm identifies distinct types of human adenovirus motion in host cells. *J. Struct. Biol.* 159(3):347–358, 2007.

J. A. Helmuth, C. J. Burckhardt, U. F. Greber, and I. F. Sbalzarini. Shape reconstruction of subcellular structures from live cell fluorescence microscopy images. *J. Struct. Biol.* 167(1):1–10, 2009.

J. A. Helmuth and I. F. Sbalzarini. Deconvolving active contours for fluorescence microscopy images. In *Proc. Intl. Symp. Visual Computing (ISVC)*, volume 5875 of Lecture Notes in Computer Science, pages 544–553, Las Vegas, USA, November 2009. Springer.

J. A. Helmuth, G. Paul, and I. F. Sbalzarini. Beyond colocalization: inferring spatial interactions between subcellular structures from microscopy images. *BMC Bioinf.* 11:372, 2010.

LANGUAGES

German (mother tongue)
 English (fluent)
 French (basics)