Diploma Thesis

Department for Theoretical Physics,

TU Berlin

Diffusion Processes with Hidden States from Single Molecule Experiments

submitted by

Arash Azhand

Berlin, 3rd April 2009

1st Supervisor:

Prof. Dr. Eckehard Schöll, PHD

(Department for Theoretical Physics, TU Berlin)

2nd Supervisor:

Dr. Frank Noé

(Department for Mathematics and Computer Science, FU Berlin)

Five years ago many people saved my life
in numerous ways,
when I was nearly loosing it.

This work is dedicated to all of them.

Abstract

Anomalous diffusion is a ubiquitous phenomenon in biophysical processes such as the motion of Biomolecules within cellular compartments. Often, the apparently complex diffusion behavior arises from the fact that the observed molecule switches between different states, in each of which the diffusion is simple but different. Modern imaging techniques, such as single-molecule fluorescence particle tracking allow to track the diffusion behavior down to individual molecules by measuring the time-dependent position of the molecule, but generally do not reveal the state each molecule is in. Understanding the physics behind such experimental data amounts to solving an inverse problem uncovering the set of hidden molecular states, their local diffusion behavior, and their transition rates. A computational method for solving this inverse problem based on Hidden Markov models is presented in this work.

Vertebrate vision signal transduction, namely the translation of light signals, which enters the eye in form of single photons, to an electrical signal that arrives at the visual cortex, is one of the most-studied processes in biophysics. The primary vision transduction taking place in the retinal rod cells responsible for dark vision has received special attention, as it is able to transform the arrival of a single photon into a well-defined and reproducible electric signal.

Within these rod cells are stacked approximately 1000 "disc membranes", each of which have a set of different proteins bound to them that are essential to rod cell vision transduction. In the initial stage, individual photons are absorbed in the retinal ligand bound to rhodopsin molecules, leading to photo-isomerization and thus rhodopsin adopting an ”active” conformation state, which stays active until it is actively shut off by a different cellular process.

In the next step, transducin proteins collect the vision signal by binding to activated rhodopsins and subsequently themselves transform to an activated state via which they activate the next protein in the cascade. Transducin plays a key role in the signal amplification, as a single active rhodopsin can activate many transducins in its lifetime. In terms of its diffusion behavior, Transducin can exist in different molecular states:

1. It can diffuse in between the discs, in a solvent which is mainly consisting of water and ions.
2. It can be bound to the disc membrane via a lipid anchor and diffuse laterally amongst the membrane lipid molecules.
3. It can form a metastable pre-complex with rhodopsin (either inactive or active), and diffuse laterally within the membrane together with the rhodopsin it is bound to.

Each of these states have different diffusivities and can inter-convert, leading to a complex overall diffusion behavior. If activated rhodopsins are available, state 3 can transform into a tightly bound active complex and transducin is activated. However, all states $1-3$ are available in the dark (without activated rhodopsin). The research group of Prof. Dr. Ulrike Alexiev at the Free University of Berlin performs single molecule fluorescence tracking experiments on labeled transducin

molecules from bovine retinal rod cells. In these experiments the individual transducin positions on the disc membrane were tracked in time and corresponding two-dimensional trajectories have been collected. The challenge addressed in the present work is to derive a computational method which can detect the hidden states $1-3$ from the data and estimate the local diffusion laws in each state.

In order to model this hidden states sequences, the mathematical concept of Hidden Markov models is used. Within this thesis a Markov chain with three states is assumed for modeling the three molecular states given above. These states interconvert with fixed rate constants or switching probability, modeling the binding/unbinding of transducin to/from the membrane and rhodopsin. Within the states, a local Langevin equation is active which governs the diffusion behavior at the time. Methods to estimate the most likely Hidden Markov model under the given observation are presented.

The procedure proposed here is demonstrated to work reliably on model data. The available experimental data was found to be insufficient in order to parametrize state 3. For state 1 and 2, diffusion models could be obtained from the particle tracking trajectories with diffusivities that agreed to complementary experimental data.

Zusammenfassung[1]

Anomale Diffusion ist ein universelles Phänomen in biophysikalischen Prozessen wie zum Beispiel der Bewegung von Biomolekülen innerhalb der Zelle oder in den Zellkompartimenten. Oft liegt das scheinbar komplexe Diffusionsverhalten darin begründet, dass das beobachtete Molekül zwischen verschiedenen Zuständen umschaltet, wobei in jedem dieser Zustände die Diffusion einfach aber verschieden ist. Moderne Abbildungsverfahren, wie zum Beispiel die Einzelmolekülverfolgung mittels Fluoreszenz, erlauben die Positionsmessung einzelner Moleküle, welche diffundieren, in der Zeit. Jedoch nicht die einzelnen Zustände, in denen sich das betreffende Molekül befindet. Die Physik hinter solchen experimentellen Daten zu verstehen führt auf das Lösen eines inversen Problems hinaus. Dies soll die Menge der versteckten Zustände aufdecken, die einzelnen Diffusionsverhalten in den Zuständen und die Übergangsraten zwischen den Zuständen. Eine rechnerbetonnte Methode zur Lösung dieses inversen Problems, basierend auf ein "Hidden Markov"-Modell (versteckte Markovzustände) wird in dieser Arbeit präsentiert.

Die visuelle Signaltransduktion bei den Wirbeltieren, das heißt die Übersetzung beziehungsweise die Transformation der Lichtinformation, die in Form von einzelnen Lichtphotonen in das Auge gelangt und zu einem elektrischen Signal transformiert wird, das schließlich von der Netzhaut ihren Weg durch das Gehirn in das visuelle Cortex findet, ist eines der am meisten untersuchten Prozesse in der Biophysik. Die primäre visuelle Transduktion findet in den über 120 Millionen Stäbchenzellen statt, welche über die ganze Netzhaut verteilt sind und für das sogenannte Dunkel-Sehen bei den Wirbeltieren (zu denen auch der Mensch gehört) verantwortlich sind. Innerhalb dieser Stäbchenzellen sind etwa 1000 "Disc-Membrane" übereinander gestapelt, einzelne durch Doppellipidschichten von der intrazellulären Umgebung getrennte Vesikeln.

Diese beherbergen als integrale Bestandteile ihrer Membran unter anderem die Rhodopsin- Moleküle, welche das Licht absorbieren und dadurch eine Konformationsänderung erfahren und aktiv bleiben, bis sie durch ein weiteres Protein deaktiviert werden. Darüber hinaus diffundieren zwischen den Disc-Membranen und auf der Membran Transducin-Moleküle, die nach Kollision mit Rhodopsin mit einer festen Bindungsrate an Rhodopsin binden und mit diesem einen Komplex formen. Rhodopsin aktiviert wiederum dadurch das Transducin-Molekül, während Transducin wiederum das nächste Protein in der Kaskade aktiviert. Transducin spielt eine Schlüsselrolle in der Signalverstärkung, da ein einzelnes Rhodopsin während ihrer aktiven Lebenszeit in der Lage ist, mehrere Transducin-Moleküle zu aktivieren. Bezüglich des Diffusionsverhaltens kann das Transducin-Molekül in drei verschiedenen molekularen Zuständen existieren:

1. Die Bewegung des Transducins in der Zellflüßigkeit zwischen den Disc-Membranen,
2. die Bewegung des Transducins auf der Membran und
3. es kann mit Rhodopsin einen metastabilen Prä-Komplex (aktiv oder inaktiv) formen und dann als Komplex lateral auf der Membran diffundieren.

[1] Recapitulation of the thesis in German.

Jede einzelne dieser Zustände besitzt ein eigenes Diffusionsverhalten, wobei zwischen den Zuständen Übergänge möglich sind. Wenn aktiviertes Rhodopsin vorhanden ist, kann der dritte Zustand zu einem fest gebundenen Komplex von aktiviertem Rhodopsin und Transducin transformieren. Jedoch existieren alle drei Zustände auch während der sogenannten "Dunkel-Phase", wenn Rhodopsin nicht aktiviert ist. Die Arbeitsgruppe von Prof. Dr. Ulrike Alexiev an der Freien Universität Berlin hat Einzelmolekül-Fluoreszenz-Experimente an gemarkerten Transducin-Molekülen durchgeführt. Im Rahmen dieser Experimente wurden einzelne Transducine zeitlich verfolgt und somit wurden einzelne zweidimensionale Trajektorien aufgenommen. Dabei ist zu erwähnen, dass die entsprechenden Stäbchenzellenmembrane und die gelabelten Transducine aus Augen von Kühen gewonnen und für die Untersuchungen im Rahmen der Einzelmolekül-Verfolgungs-Experimente präpariert wurden.

Die Herausforderung in dieser Arbeit besteht darin, eine rechnergestützte Methode zu entwickeln, die in der Lage ist, die versteckten Zustände $1-3$ aus den Beobachtungsdaten zu schätzen und darüberhinaus die lokalen Diffusionsgesetze innerhalb der einzelnen Zustände. Um diese Abfolgen versteckter Zustände zu modellieren, wird das sogenannte mathematische Konzept der versteckten Markov-Ketten herangezogen. Im Rahmen dieser Arbeit wurde eine Markov-Kette mit drei Zuständen angenommen. Die Annahme von drei Zuständen folgt aus den oben aufgeführten drei zu erwartenden Bewegungsphasen der Transducin-Moleküle, welche als metastabil angenommen werden.

Die in dieser Arbeit aufgeführte Methode zeigte sich als erfolgreich bei Anwendung auf künstliche Modelldaten. Auf Grund der nicht so großen Menge an experimentellen Daten (ca. 5700 Datenpunkten) war die Parameterschätzung basierend auf ein Modell mit drei versteckten Zuständen nur teilweise erfolgreich. Es war nicht möglich den dritten Zustand auszumachen. Dagegen waren die geschätzten Diffusionskoeffizienten für den ersten und zweiten Zustand in guter Übereinstimmung mit schon früher durch andere berechneten Transducin-Diffusionskoeffizienten.

Contents

List of Figures

1 Motivation

„NEC FASCES, NEC OPES,
SOLA ARTIS SCEPTRA PERENNANT“ [57, p. 105]

”Neither high agencies, nor power,
solely the scepter of science outlast”

(Tycho Brahe, motto, carved in stone over the north entrance
of the Stjerneborg observatory, built by Brahe in the year 1584)[1]

Within living memory or at least since humanity has emerged conscious awareness, we possess a quenchless desire to give sense to all what we percept. At all times mankind has the wish to understand its surrounding environment, and this wish is inseparably correlated with the fact that we inevitably create models when we adept a certain phenomenon in nature.

A model is in general an image of the phenomenon, an image that we make, in order to understand and maybe forecast future phenomena. Models are emerging in our brains backwards, meaning that experience is coming first and afterwards is leading to a model. Though other, earlier thought models can be the basis of our new models it is at least experience that enables us to estimate a new model.

A model in mathematical sense is a set of mathematical objects that are connected with each other by functional relationships.

A Theory is an aggregation of different models and they are described by a set of axioms that build up the basis of the theory. Propositions can be derived from these axioms within the theory, and propositions in turn can be checked by observation. While the propositions can be tracked back to the axioms of a theory, there is no proof in a mathematical sense for the axioms themselves, axioms are moreover convention.

As such, theories can not be proved at last but in the best case experience a kind of endorsement. Endorsement by numerous observations is verifying different propositions of the theory. Nevertheless a theory is never immune against falsification of at least one of it's propositions. Thus a theory can be falsified but not verified in perpetuity.

A good theory is characterized by its capability to combine and interconnect such phenomena that on first sight have not much to do with each other. By this a theory has a more universality than a

[1] German translation: „Weder hohe Ämter, noch Macht, einzig die Zepter der Wissenschaft überdauern“ - Tycho Brahe.

mere model.

In the appendix is presented an example, in order to show the extraordinary relationship between observation and theory in the natural sciences on the one hand and the interrelation between models and theories on the other hand (see section A.1 on page 99).

We see here an example for the modeling that grounds on observation (Tycho Brahe) which led to the laws of planetary motion (Johannes Kepler) and later to the formulation of probably the first scientific theory of mankind by Isaac Newton, from which earlier formulated models such as the planetary motion model of Kepler and others can be derived.

The motivation of this thesis is the decent ambition to take into account some observation and on the basis of this observation to set up a model that hopefully broadens our view of the considered phenomenon. Maybe this will help us in the future to understand the whole process in the sense of something like a theory, although the process herein is so complex that we can not possess the certainty to reach this ambition but hope will not leave us.

The phenomenon in consideration is the first stage of the process which is taking place in the eye and converts the light into an electrical signal. It is nothing less than the process which finally after numerous steps leads to vision.

When it was pointed out at the beginning of this motivation chapter that we own the desire to understand the world that surrounds us, then vision is maybe the most important gift that we have received from nature, in order to reach this understanding. It is our vision that let us see the world before interpretation and model building, based on what we see, is taking place in our brains. For this reason vision is the gate to our consciousness, as anatomically the retina, where the conversion of light information into electrical information is taking place, is the gate to our brain.

The history of theories on the visual sense is dating back from the Greek philosophers like Plato and Aristotle, and it was finally Kepler who, grounding on the works of middle ages Arabian scientists, established a model for the process of seeing which is the basis for our today's understanding of the visual sense. In his work "Dioptrice" from 1611 Kepler constituted optics as an area of science. In this and in earlier works he condemned ancient associations that described vision through the thesis that rays extinguish from the eye and comprehend the considered object. Moreover light rays extinguish from every point of an object into all directions of space and reach for instance also through the pupil the inner eye.

Today it is of great importance and motivation to understand vision even deeper, to understand the single sub elements, which are finally leading to visual sense. To begin this ambition, an introduction will be given to the visual transduction process as a whole and the first stage of this process in particular in the next chapter.

For modeling this and other processes in biology, we have a priceless advantage compared to our scientific ancestors. Today we possess both enormous experimental and theoretical tools enabling us to reach such heights of knowledge about nature that our ancestors may never have been dreamed of.

Experimentally we are able to make visible such events and processes that are laying in the tiny world of the microcosm down to the range of nanometers, and on the other side we have telescopes, enabling us to see events and processes up to 10 billion light years away from the earth.

Theoretically we profit from enormous developments of the last two and three centuries, considered both physics and mathematics. In physics, especially the developments related to the theory of quantum mechanics changed our view on the world of tiny atoms and molecules. The objects, with which quantum mechanics as a probabilistic theory is dealing, are not much different from the objects which for instance the classical statistical mechanics is dealing with. Both theories do not give deterministic predictions on the dynamics of single events but moreover on the dynamics of ensembles of similar events.

The theory of statistical mechanics will play a significant role in this thesis, since the physical processes we aim to model can be displayed by mathematical objects that are elements of both statistical physics and stochastic mathematics.

The third and not less important advantage that we own compared to our ancestors in the history of natural sciences, is the computer as a barely replaceable tool for modeling. Complex physical and biological processes, like the one presented in this thesis, can not be investigated as thoroughly as we wish to, by means of analytical methods, merely by pencil and paper. Through their intrinsic stochastic nature, they can be represented for instance through numerous equivalent models nearly equal likely. That is why the computer will be the tool for searching the mathematical model spaces, to compare the models with each other and discriminate them by their capability to explain the observation.

Thus to conclude this introductory chapter, the leitmotif of this thesis will be the unification of theory, experiment and computation, with the purpose to better understand the nature of a special biological process, that is the process of vision.

2 Visual Transduction

”The eye owes its existence the light.
From indifferent animal auxiliary organs
the light arouses for itself an organ,
that shall become its own kind,
and so constitutes the eye ajar the light for the light,
therewith the inner light shall counter the outer light.”
(Johann Wolfgang Goethe)[1]

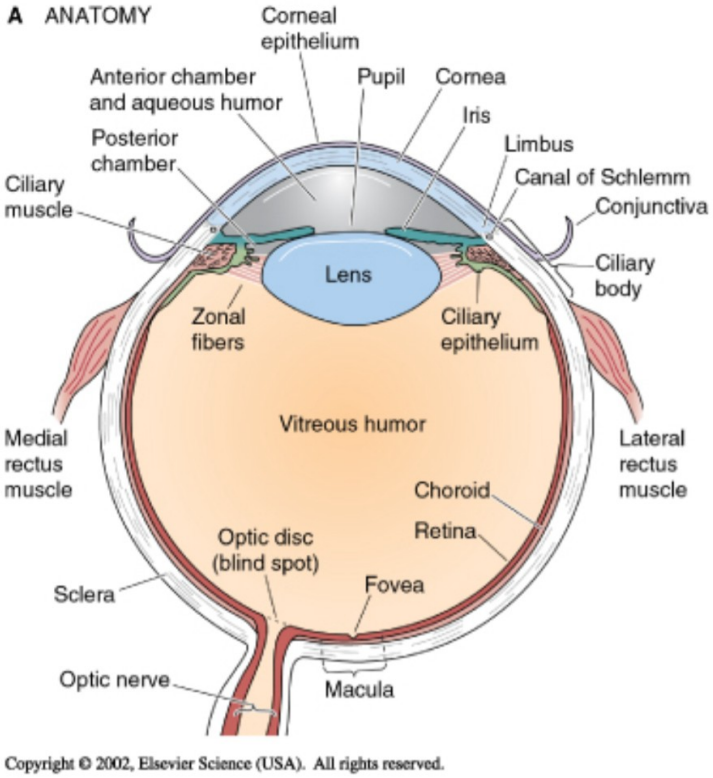


Figure 2.1: Anatomy of the eye. Light is propagating through the lens which focuses the light on the retina.

[1] German translation: ”Das Auge hat sein Dasein dem Licht zu danken. Aus gleichgültigen tierischen Hilfsorganen ruft sich das Licht ein Organ hervor, das seinesgleichen werde, und so bildet sich das Auge am Lichte fürs Licht, damit das innere Licht dem äusseren entgegentrete” (Johann Wolfgang von Goethe, Goethes Naturwissenschaftliche Schriften (Kürschner) „Entwurf einer Farbenlehre“, Bd. 3, S. 88).

Vision is a complex process which not only enables us humans to percept the environment by absorbing and converting the visual electromagnetic spectrum in between 400 *nm* and 700 *nm* which is reflected on or emitted from our surrounding, to electric potentials. Therefore we can react on and master any kind of challenge nature gives us.Visual processing which is the conversion of light to an electrical signal in retinal photo-receptors, consists of numerous sub processes [7].

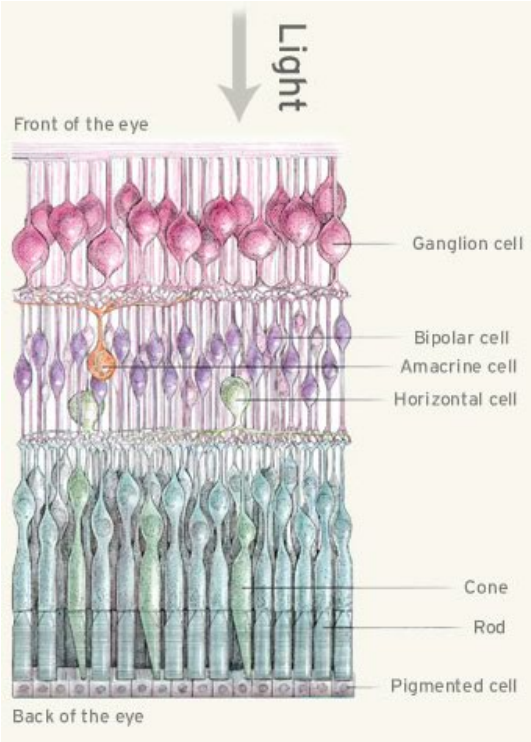


Figure 2.2: The layers of the retina. On the bottom layer the two receptor cells, Cones and Rods, are situated. Here the light, after passing through the upper layers, is converted to an electrical signal, that transmitted via the Horizontal, Amacrine and Bipolar cells, is reaching the Ganglion cells. The whole bunch of signals, gathered by these Ganglion cells will be transmitted via nerve fibers to the visual cortex of the brain.

This first stage of the visual processing will be described here. Light first propagates through the lens and will be focused on the retina (see Figure 2.1). The retina finally is the place where visual transduction, namely the conversion of light to an electrical signal takes place.The retina consists of many layers and on the inner layers the photoreceptor cells are situated (see Figure 2.2). Two types of photoreceptor cells are found there, namely about 120 million rods and about $6-7$ million cones.

The cones are necessary for color (photopic) vision and there are at least three types of cones excited by three distinct wavelength areas (red, green and blue). But here we will not discuss the cones and focus on the rods. In evolutionary terms [7] rods have arisen more recently than cones and numerically rods are the dominant class of photoreceptor cells in the human eye. Furthermore they function at exceedingly low intensities, corresponding to starlight or moonlight.

This means that rods are saturated at normal light conditions. It has been shown that rods are able to ”detect individual photons of light [Hecht et al., 1942], and the subsequent circuitry of the scotopic visual system is designed to be able to detect just a few photon hits scattered over large areas [Barlow et al., 1971]. This enables the visual system to operate in a ’photon-counting’ mode [Rose, 1948]. The enormous number of rods is employed primarily to catch just about every photon that reaches the retina” [7]. As we can see in Figure 2.3, Rods (but also Cones) consist of three primary functional parts: outer segment, inner segment and synaptic ending.

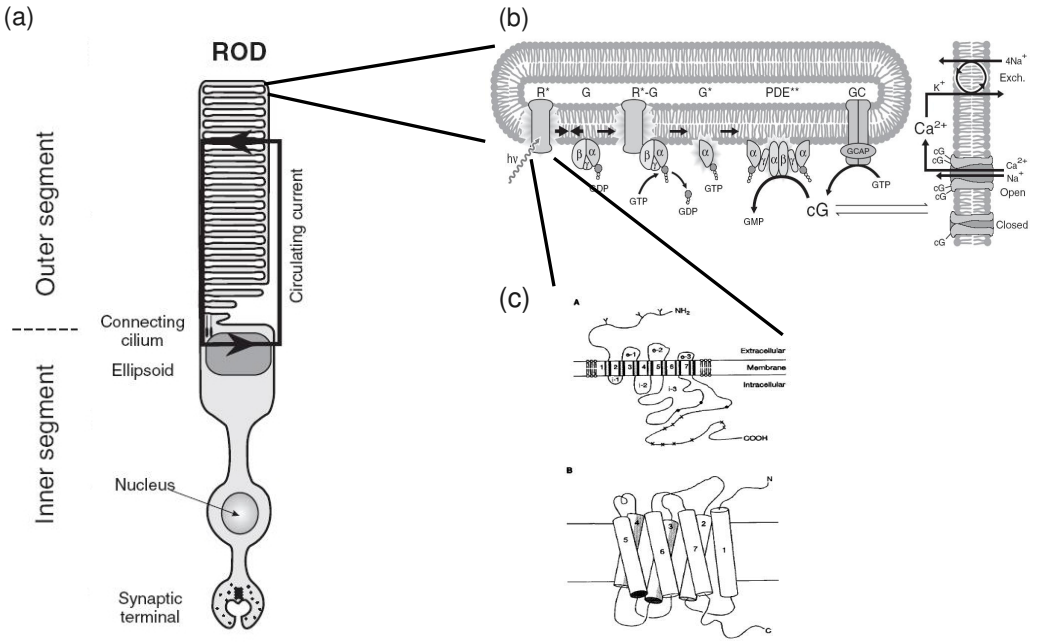


Figure 2.3: (a) The anatomy of the rod photoreceptor cell (from [7]). It is consisting of an inner segment, outer segment, a nucleus and a synaptic terminal that at the end transmits the electrical signal to the higher-ranking nerve cells. On the outer segment are situated stacks of several hundreds to thousands of vesicular discs. (b) The scheme of the visual transduction process in these retinal rod discs (presented also in Figure 2.4). (c) On the membranes of these discs the 7TM receptor molecule rhodopsin is situated (see also Figure 2.5).

For the purpose of this thesis the focus will be only on the outer segment. Several hundreds to thousands of membrane pairs are stacked along this outer segment. This structures which are completely separated from the plasma membrane, are forming intracellular organelles and are called discs. Many of the proteins and other components of the visual transduction cascade reside in (or at the surface of) this discs. The most abundant protein in the disc membrane is the light-capturing protein, called rhodopsin. The disc membranes in fact are densely packed with rhodopsin molecules, about 30.000 molecules per μm^2.

Rhodopsin is part of the 7*TM serpentine receptor* family. Without going too much into detail, one can say that "all these receptors share a common architectural theme. Their polypeptide chain makes seven passes through the membrane (see figure 2.5)" [36, p. 169]. That is the reason, why they are called seven trans-membrane or serpentine receptors.

"The 7TM receptors share not only a common architectural theme but also a common intermembranous method of signal amplification"[36, p. 169]. This mechanism grounds on the lateral mobility of the so called G-Proteins. G-Proteins are members of a large class of guanine nucleotide binding proteins and they function more or less as biochemical switches, which can turn on and turn off the activity of other molecules. In the case of the visual transduction process in rod discs the G-Protein is called *transducin*. For more information about the serpentine-receptor family and G-Proteins in general, see [36, chapter 8]. In the following we solely concentrate on

the special case of this general signal transduction process, which leads to the conversion of single photon signals to amplified electrical signals. This special case will be denoted by the term *Rhodopsin-Transducin Process*. The molecular reactions underlying the activation stages of visual transduction in retinal rod discs are described below [7] and schematically shown in Figure 2.4. For a more detailed description of the whole process see [7].

1. After the absorption of a photon of light, the rhodopsin molecule is transformed into an enzymatic active state (R*).
   Absorption of a photon by the bent 11-cis retinal isomer within rhodopsin has a high probability ($\sim 0,67$) of triggering the cis to trans isomerization of the retinoid (see figure 2.6) . Apparently, the straighter all-trans isomer no longer fits neatly into the hydrophobic pocket so that the rhodopsin molecule is strained internally, undergoing a series of molecular rearrangements that lead to its activation (as R*).
2. R* in turn catalyzes the activation of the G-protein transducin (to G*).
   Upon binding of the transducin G to R*, the Gα subunit releases its bound molecule of guanosine diphosphate (GDP) and takes up a molecule of guanosine triphosphate (GTP) from the cytoplasm (see Figure 2.4). The GTP-bound form of Gα represents the active entity, denoted as G*, which carries forward the activation signal. The activated rhodopsin R* separates from the activated transducin, and unaltered by the interaction, is free to interact with additional molecules of the G-protein, thereby catalyzing the activation of many G* during its active lifetime.
3. Transducin again activates the effector protein phosphodiesterase (PDE) to PDE*.

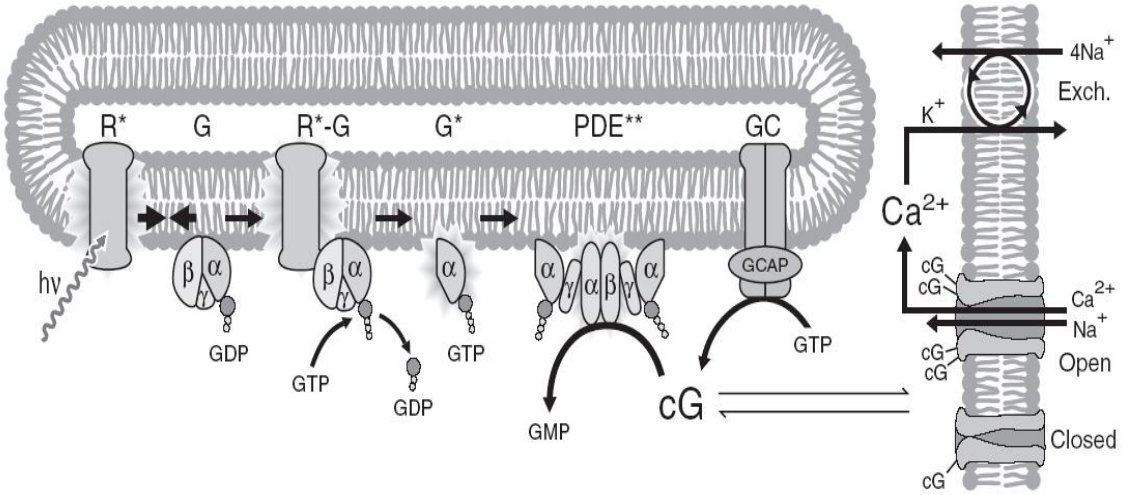


Figure 2.4: Scheme of the visual transduction process in the retinal rod discs (from [7]). For details see text. The scheme of the $7TM-$ or serpentine-receptor rhodopsin (R or its activated form R^*) is shown in Figure 2.5.

4. The activated PDE hydrolyzes the diffusive messenger cyclic guanosine mono-phosphate (cGMP) so that the cytoplasmic concentration of cGMP decreases, leading to closure of cGMP-gated ion channels in the plasma membrane. Closure of these channels has the dual

effect of generating the photoreceptor's electrical response (by reducing the circulating current and hyper-polarizing the membrane voltage), and of causing a reduction in cytoplasmic calcium concentration.

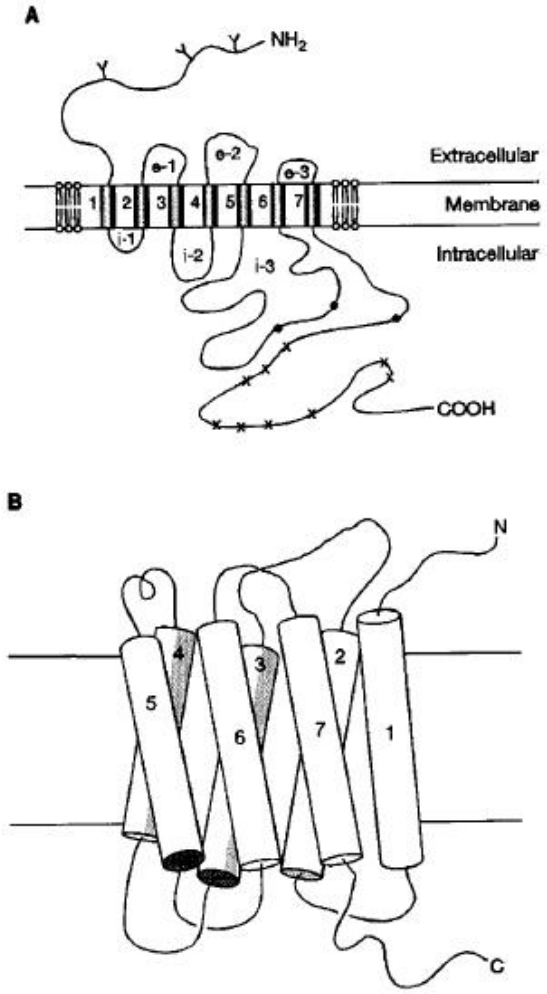


Figure 2.5: Scheme of the 7TM serpentine receptor (from [36, p.170 , Fig. 8.3]). (A) The seven transmembrane helices are shown as columns in the membrane, labelled 1–7. The N-terminal sequence is extracellular and is usually glycosylated (Ys). The extracellular loops are labelled $e-1, e-2, e-3$ and may also sometimes be glycosylated. The intracellular loops $i-1, i-2, i-3$ provide recognition surfaces for designated G-proteins. The black spots represent phosphorylation sites for PKA and the crosses represent sites for specific desensitising enzymes (e.g. b-ARK in the case of the b-adrenergic receptor). (B) 'Barrel of staves' three-dimensional conformation of a 7TM receptor in a membrane.

Finally when the membrane is hyperpolarized, the membrane potential equilibrate to its ground state and an electrical signal will be transmitted via the Horizontal, Amacrine and Bipolar cells, and finally is reaching the Ganglion cells. The whole bunch of signals, gathered by these Ganglion

cells will be transmitted via nerve fibers to the visual cortex of the brain (see also Figure 2.2).

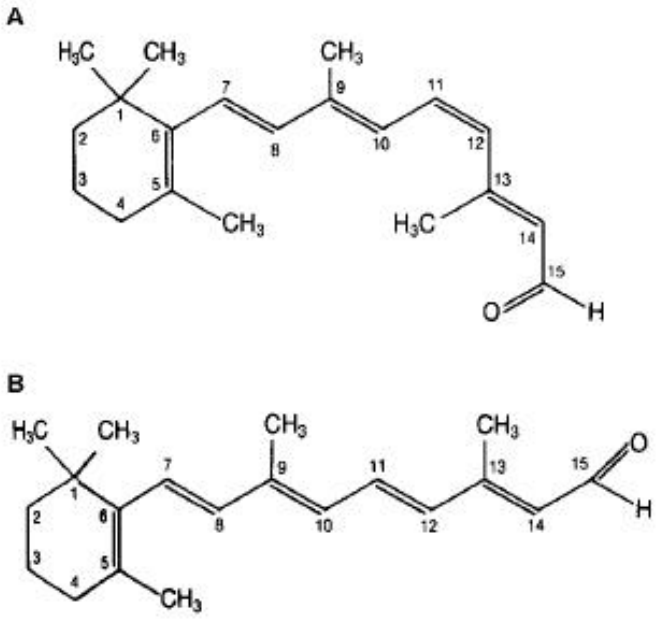


Figure 2.6: (A) 11-cis-retinal and (B) all-trans-retinal. *C* : Carbon, *H* : Hydrogen, *O* :Oxygen, *N* : Nitrogen. The numbers are atom positions. (from [36, p. 190, Fig. 8.23])

3 Theory

"The theory is the net, that we throw out,
in order to capture the world, to rationalize it,
to explain and to control it.
We work on making the stitches of the net narrower."
(Karl Popper) [1]

Now that the process which has to be modeled is introduced we have to present the basic theory, on which modeling of the associated process will take place. Biological systems like visual processing for instance are intrinsically stochastic in their nature.

The physical theory that is the framework for such systems is the theory of thermodynamics.

> "Thermodynamics is mainly concerned with the transformations of heat into mechanical work and the opposite transformations of mechanical work into heat" (Enrico Fermi, [20]).

This quote from a book about thermodynamics written by Enrico Fermi specifies the case very well. At the physiological temperature of $37^{\circ}C$ heat is the driving force and the molecules within the cell are defenseless against it.

Due to their low mass they librate and pace around while crashing permanently with molecules of water and with other molecules. Through this the motion of single molecules in a biological environment has an intrinsically random nature. That is why mathematical modeling of molecular dynamics also will be done in the framework of stochastic processes.

In the next section we will introduce the notion of inverse problems as the gate to the further theoretical considerations that are needed to model the first stage of visual transduction process on the ground of single transducin molecule trajectories which have been tracked experimentally.

[1] German translation: "Die Theorie ist das Netz, das wir auswerfen, um 'die Welt' einzufangen - sie zu rationalisieren, zu erklären und zu beherrschen. Wir arbeiten daran, die Maschen des Netzes immer enger zu machen" (Karl Popper, Logik der Forschung).

3.1 Inverse Problems and Bayes Theorem

Let there be a particular type of observation denoted by O and a set of arbitrary parameters λ denoting a model. Then it is assumed that there exists a function G relating λ and O with one another in the following way:

$$G(\lambda) = O. \tag{3.1}$$

The *"forward problem"* is to find O when λ and an arbitrary initial condition O_0 are given.

In the case of stochastic processes both λ and O may be noise-driven. Due to the stochasticity of the model the forward problem surely does not lead always to the same observation for a given initial condition, but rather to a distribution of observations.

But in many cases, as also in this thesis, the task is to find the *appropriate* model λ for any given observation O. This is the so called *"inverse problem"*.

There are numerous examples for inverse problems in the natural sciences: tomography in medicine, radio-astronomical imaging, navigation, image analysis, geophysics, and so on[2].

The solution of such inverse problems are by some reasons hard to achieve [1]:

1. *Existence.* There may be no model λ that can explain the observation.
2. *Uniqueness.* If a model λ exists, it may not be unique in the sense that there may exist many distinct models which also reproduce the observation adequately.
3. *Instability.* A small change in the observation can lead to an enormous change in the estimated model. In this case one speaks of *"ill-posed" problems*.

The more interested readers are referred to [1] and especially to [38].

One basic tool to achieve model estimation on inverse problems is the so called "Bayes theorem":

Theorem 3.1 (Bayes theorem).

Let n be a positive integer, $\{\lambda_n\}$ a set of parameter tuples and O the given observation, thus we have:

$$P(\lambda_n|O) = \frac{P(\lambda_n)P(O|\lambda_n)}{P(O)}, \tag{3.2}$$

with $P(O) = \sum_n P(\lambda_n)P(O|\lambda_n) > 0$.

The Bayes theorem, published in 1763, is supposed to yield an *"inverse probability"*. Simplifying the case, we can also write

$$P(\lambda|O) \propto P(\lambda)P(O|\lambda), \tag{3.3}$$

[2]There exists a journal on the purpose, reachable through the address: http://www.iop.org/EJ/journal/IP

where $P(\lambda)$ is the a priori probability (prior), $P(\lambda|O)$ the a posteriori probability (posterior) of the model λ. $P(O|\lambda)$ is the so called Likelihood $\mathscr{L}$ (see section 3.4 on page 21), which in fact is a measure for the fitness of an arbitrary model to the observation.

For an interesting article on Bayesian networks in computational biology see [45].

3.2 Markov Chains, Markov Processes and Markov Property

It is unlikely that the Russian mathematician A. A. Markov during his lifetime could have an idea of the enormous influence which his work in general, and especially his work on those stochastic processes that carry his name had and until today has on so numerous fields of sciences. The consideration on Markov processes, and everything what has in any way to do with it is something which plays an important role in modern sciences, beginning with pure and applied mathematics and going on with theoretical physics, chemistry and biology.

Those, who are interested on the life and work of A. A. Markov, are referred to [3]. In the abstract Basharin et al. for example are pointing out:

> "The Russian mathematician A. A. Markov (1856-1922) is known for his work in number theory, analysis, and probability theory. He extended the weak law of large numbers and the central limit theorem to certain sequences of dependent random variables, forming special classes of what are known as Markov chains. For illustrative purposes Markov applied his chains to the distribution of vowels and consonants in A. S. Pushkin's poem *Eugeny Onegin*."

At the beginning of our considerations on Markov processes a definition is given which on the one hand show the simplicity of the topic, and on the other hand maybe can explain, why it is so easy to model real life systems via Markov processes.

Definition 3.1 (Markov Process and Markov property).

Let $\{X(t)\,|\,t \in T\}$ be a stochastic process defined on the parameter set T, with values defined on the state space S. This process is called Markov process if it satisfies the following property (Markov property):

$$\begin{aligned} & P[X(t_n) = x_n \,|\, X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \cdots, X(t_0) = x_0] \\ = \;& P[X(t_n) = x_n \,|\, X(t_{n-1}) = x_{n-1}], \end{aligned} \tag{3.4}$$

$$t_0 < t_1 < t_2 < \cdots < t_{n-2} < t_{n-1} < t_n$$

The assertion of the above definition (3.1) is that the state of a particular process at any point in time (provided that the parameter set T is denoting the time) depends only on the state of the process at the previous point in time. In other words: the current state includes the whole information one needs, in order to determine the future distribution of the process.

The reason why we have mentioned that it would be easy to model real life systems via Markov processes, is now clear: a stochastic process without memory is maybe one of the simplest one can formulate.

According to [25] and [12, 13] we will classify those processes with discrete state spaces S and parameter sets T as "Markov chains", and those processes with continuous S and T as "Markov processes". First Markov chains will be discussed, afterwards Markov processes. Thus we can define Markov chains as follows:

Definition 3.2 (Markov chains).

A Markov chain is a set of discrete valued random variables that satisfy the following property:

$$\begin{aligned} & P[X_n = i_n \,|\, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \cdots, X_1 = i_1, X_0 = i_0] \\ = \;& P[X_n = i_n \,|\, X_{n-1} = i_{n-1}]. \end{aligned} \tag{3.5}$$

For a schematic representation of simple Markov chain with three states $s_1, s_2, s_3 \in S$ see Figure 3.1 in the next page.

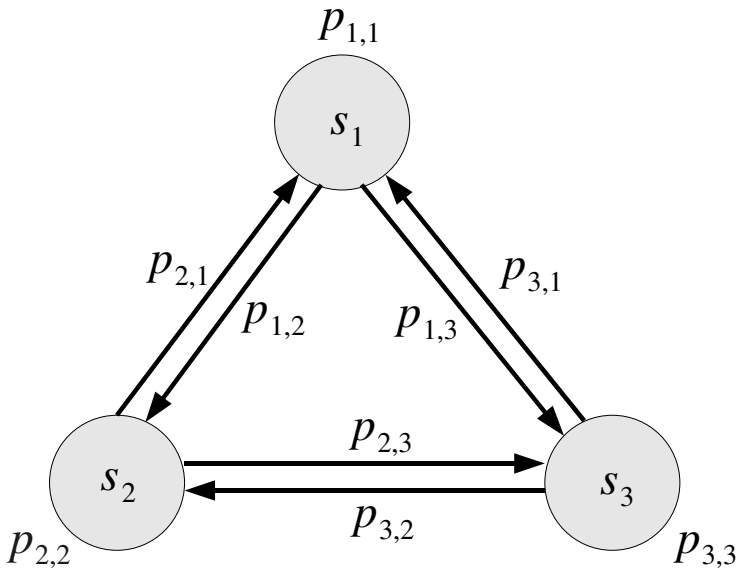


Figure 3.1: Scheme of a Markov chain. Schematic representation of a simple Markov chain with three states $s_1, s_2, s_3 \in S$. The $p_{i,j}$ with $i, j \in \{1,2,3\}$ are transition probabilities between states i and j.

From now on we are able to derive all properties of Markov chains (see [25, p. 333]).

Let the state space of the chain S be the set of non-negative integers: $0,1,2,3,4,\ldots$, furthermore one can define the probability $P[X_n = j \,|\, X_{n-1} = i]$ as the transition probability of state i to state j. In this case the chain is called *time-homogeneous*, if

$$P[X_n = j \,|\, X_{n-1} = i] = P[X_{n+m} = j \,|\, X_{(n+m)-1} = i],$$

$$n = 1,2,\ldots,\ m \geq 0,\ i,j \in S. \tag{3.6}$$

Thus we can write

$$p_{i,j} \equiv P[X_n = j \,|\, X_{n-1} = i], \tag{3.7}$$

and define the matrix of transition probabilities by

$$\mathscr{T} = \begin{pmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,j} & \cdots \\ p_{1,0} & p_{1,1} & \cdots & p_{1,j} & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \\ p_{i,0} & p_{i,1} & \cdots & p_{i,j} & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \end{pmatrix}. \tag{3.8}$$

We will write $i \mapsto j$ if $p_{i,j} > 0$, which means that the chain can jump directly from i to j.

We can think of the operation of the chain as follows: The chain starts at time 0 in some state, say $i_0 \in S$. At the next time step the chain jumps to a neighboring state i_1 with the probability p_{i_0,i_1}, provided that $i_0 \mapsto i_1$. It may be the case that this jump is immediately back to the state itself, that is $i_0 = i_1$. We call such an occurrence a *self-loop*. This procedure is repeated so that at step n the chain is in some state i_n, where

$$i_0 \mapsto i_1 \mapsto i_2 \mapsto \cdots \mapsto i_{n-1} \mapsto i_n. \tag{3.9}$$

A sequence of states with (3.9) is called a *path*.

Given an initial state, there is a set of possible paths that can be taken by the Markov chain. This is called the set of sample paths. One particular sample path, taken by the chain is denoted by

$$i_0, i_1, i_2, \ldots i_{k-1}, i_k, \ldots . \tag{3.10}$$

A graphical representation for such a sample path (sample sequence) with N elements, based on a simple three-states Markov chain, as shown in Figure 3.1, is given by Figure 3.2.

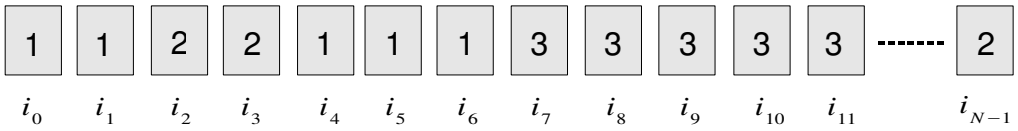


Figure 3.2: Schematic representation of a Markov chain path (sequence), based on a three state Markov chain, shown by Figure 3.1, with states $1,2,3$ and N elements $i_0, i_1, i_2, \ldots, i_{N-1}$.

We furthermore will write $i \overset{n}{\rightsquigarrow} j$ if there exist a path of n steps between i and j.

The Markov property (see definition (3.2), equation 3.5) allows us yet to write down an expression for the total probability that a particular sample path is taken by the chain by multiplying the transition probabilities for each step of the sample path. In particular, the probability of sample path (3.9), (3.10) is given by

$$p_{i_0,i_1} \cdot p_{i_1,i_2} \cdot p_{i_2,i_3} \cdot \cdots \cdot p_{i_{n-2},i_{n-1}} \cdot p_{i_{n-1},i_n}.$$

It is clear from the law of total probability that the overall probability for the chain to be in state j after n steps, is equal to the sum of the probabilities of all n-step sample paths from state i to state j. For $n = 1$ this probability is simply $p_{i,j}$, and is given by the entry in position (i, j) of transition matrix $\mathscr{T}$ (3.8). For $n = 2$ the probability of all sample paths $i \overset{2}{\rightsquigarrow} j$ can be written as

$$\sum_{k \in S} p_{i,k} p_{k,j}, \quad k \in \{0,1,2,\ldots,n\}$$

which is simply the entry in position (i, j) of the transition matrix $(\mathscr{T})^2$.

Now let $p^n_{i,j}$ denote the (i, j)-entry of the matrix $\mathscr{T}^n$ (we define $\mathscr{T}^0 = \mathscr{I}$, with $\mathscr{I}$ denoting the identity matrix). Then it is easy to see that $p^n_{i,j}$ is the probability that the process is in the state j, n steps after being in state i. The rules for matrix multiplication imply that

$$\mathscr{T}^n = \mathscr{T}^m \mathscr{T}^{n-m}, \quad 0 \leq m \leq n, \tag{3.11}$$

or written in terms of the matrix entries:

$$p^n_{i,j} = \sum_{k \in S} p^m_{i,k} p^{n-m}_{k,j}, \quad 0 \leq m \leq n. \tag{3.12}$$

Equations like (3.11) and (3.12) are called the *Chapman-Kolmogorov equations*.

[25] introduces also a very elegant "state-probability-vector" formalism which we want to use here for our purpose. So let us define the vector

$$\vec{\pi} \equiv (\pi_0^0, \pi_1^0, \pi_2^0, \ldots),$$

where π_i^0 is the probability that the initial state is i at time 0. Thus the probability that the process is in state j at step 1 is given by

$$\pi_j^1 = \sum_{k=0}^{\infty} \pi_k^0 p_{k,j}, \quad j = 0,1,2,\ldots. \tag{3.13}$$

In the same manner one can define the state-probability-vector for the nth step of the chain

$$\vec{\pi}^n \equiv (\pi_0^n, \pi_1^n, \pi_2^n, \ldots),$$

and thereupon is able to write (3.13) in a simpler form:

$$\vec{\pi}^1 = \vec{\pi}^0 \mathscr{T}.$$

The Markov property enables us now to consider $\vec{\pi}^1$ as initial probabilities for the next step of the one-step Markov chain and thus we can write

$$\vec{\pi}^2 = \vec{\pi}^1 \mathscr{T}.$$

More generally it is

$$\vec{\pi}^n = \vec{\pi}^{n-1} \mathscr{T}, \quad n = 1,2,\ldots \tag{3.14}$$

Iteration of this recurrence leads finally in a natural way to

$$\vec{\pi}^n = \vec{\pi}^0 \mathscr{T}^n. \tag{3.15}$$

Thus the conclusion of this equation (3.15) is the fact that the probability of the chain to be in the state j after n steps is just the sum of the probabilities along all sample paths $i \overset{n}{\rightsquigarrow} j$ weighted by the probability of starting in state i. In other words:

Dependent on starting in state i, the value of $\vec{\pi}^n$ is simply the ith row of $\mathscr{T}^n$, that is

$$(p_{i,0}^n, p_{i,1}^n, \ldots) = \underbrace{(0,\ldots,0,1,0,\ldots 0)}_{1 \text{ in } i\text{th position}} \begin{pmatrix} p_{0,0}^n & p_{0,1}^n & \cdots & p_{0,j}^n & \cdots \\ p_{1,0}^n & p_{1,1}^n & \cdots & p_{1,j}^n & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \\ p_{i,0}^n & p_{i,1}^n & \cdots & p_{i,j}^n & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \end{pmatrix}, \tag{3.16}$$

while $p_{i,j}^n$ is given by the *Chapman-Kolmogorov equation*, (3.11) and (3.12).

To expand the whole case to Markov processes, which have continuous state spaces S and time parameter sets T, one has to consider instead of a transition matrix $\mathscr{T}$ the so called generator Matrix $\mathscr{Q}$, [25]. But before talking about generators, we want to make a preliminary treatment of Markov processes.

The following approach is done in order to adopt an analogous formulation of the problem, as used in the case of Markov chains [2],[25].

Where we have written $p_{i,j}^n$ for the n-step transition from the state i to state j ($i \overset{n}{\rightsquigarrow} j$), we now write $p_{i,j}(s,t)$ for the conditional probability of finding the process in the state j at time t, provided that it was in state i at time s. With the properties $s,t \in T$ and $s < t$. Thus we get:

$$p_{i,j}(s,t) = P[X(t) = j \,|\, X(s) = i]. \tag{3.17}$$

Going further and considering three different points in time $s,t,u \in T$ with the property $s < t < u$, we would find the process in state i at time s and in state k at time u (see Figure 3.3).

Immediately it is clear that the according process would have been in some intermediate states j at intermediate time points t, during its propagation from s to u. For a process, which is obeying the Markov property (3.4), the probability of this particular path is getting $p_{i,j}(s,t)p_{j,k}(t,u)$, while the Markovianity of the process entails the independence of the two-component probabilities.

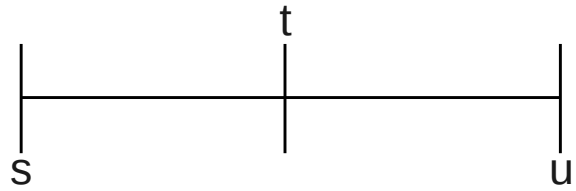


Figure 3.3: Time-Interval.

Thus taking into account any arbitrary time-interval $[s,u]$ with intermediate times $t \in [s,u]$ and $t \neq s,u$, the Markov property of the according process leads us in a natural way to the Chapman-Kolmogorov equation

$$p_{i,k}(s,u) = \sum_j p_{i,j}(s,t)p_{j,k}(t,u), \tag{3.18}$$

just as it was the case for the Markov chains (see equations (3.11) and (3.12)).

After this rather short introduction in Markov processes, we should come to the generator matrix of a Markov process $\mathscr{Q}$, [25, p. 356].

The generator matrix of a Markov process has entries that are the rates at which the process jumps from state to state, these are defined by

$$q_{i,j} = \lim_{\tau \to 0} \frac{P[X(t+\tau) = j \,|\, X(t) = i]}{\tau}, \;\; i \neq j. \tag{3.19}$$

In addition we assume here that the Markov process should be time homogeneous, therefore (3.19) is independent of the time t. The total rate out of state i is denoted by q_i and is given by

$$q_i = \sum_{j \neq i}^{\infty} q_{i,j} \,. \tag{3.20}$$

The holding time of state i is exponentially distributed with rate q_i, and by definition we set the diagonal entries of $\mathscr{Q}$ equal to minus the total rate

$$q_{i,i} = -q_i \,. \tag{3.21}$$

This implies that the row sums of matrix $\mathscr{Q}$ are equal to zero. Finally the generator matrix $\mathscr{Q}$ of a Markov process is given by

$$\mathscr{Q} = \begin{pmatrix} -q_0 & q_{0,1} & \cdots & q_{0,j} & \cdots \\ q_{1,0} & -q_1 & \cdots & q_{1,j} & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \\ q_{i,0} & q_{i,1} & \cdots & -q_i & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \end{pmatrix}. \tag{3.22}$$

Thereupon one can formulate the stationary probabilities of a Markov process as a set of linear equations that are expressed in terms of the generator matrix. Without going too much into detail, we solely denote the result of the derivation of the *"balance equations"*. When interested in the whole derivation, one is referred to [25, p. 357]. So in the end one can see that the stationary probabilities $\vec{\pi}$ of a Markov process satisfy

$$\vec{\pi}\mathscr{Q} = 0, \tag{3.23}$$

with the additional normalization requirement that $|\vec{\pi}| = 1$. As already pointed out, (3.23) are the so called balance equations, where probability flux into a state is equated by probability flux out of a state.

3.3 Hidden Markov Models

A Hidden Markov Model (HMM, [29], [5, p. 7] and [27]) is a stochastic model in which the system being modeled is assumed to be a Markov process, or a Markov chain with unknown - thus hidden - parameters, and the goal is to determine these hidden part of the system from the observable path of the process. As we have illustrated in the previous section in a regular Markov

model the state is directly observable, and therefore the transition probabilities between states are the only parameters of interest. In a HMM the state is not directly observable, hence hidden. Instead there exist arbitrary variables which are influenced by these hidden states in an unknown way, thus observing and surveying the output of these variables over time allows us to estimate the sequence of hidden states behind the observation (see Figure 3.4).

Let us define a Markov chain $\{q_t \,|\, t \in T\}$ satisfying the Markov property (3.5) and being time-homogeneous, hence we can set up a transition matrix A with entries $a_{i,j} = P[q_{t+1} \,|\, q_t]$, denoting the conditional probabilities that the Markov chain will be in state j, conditioned that it previously was in state i. We will call the Markov chain sequence $Q = \{q_0, q_1, q_2, \ldots\}$ of an arbitrary length, built upon the Markov chain process $\{q_t \,|\, t \in T\}$, the sequence of hidden states. In addition let $\{o_t \,|\, t \in T\}$ be either a discrete or a continuous stochastic process, here representing the observation path. In the discrete case the probability of observing the sequence $O = \{o(0), o(1), o(2), \ldots, o(T)\}$ of the same length as the hidden sequence Q is given by

$$P(O) = \sum_{Q} P(O|Q)P(Q), \tag{3.24}$$

where the sum runs over all possible hidden state sequences $Q = \{q_0, q_1, q_2, \ldots, q(T)\}$. In the continuous case we would have continuous observation paths O and the probabilities $P(O|Q)$ would denote probability density functions (pdf) in the observation space.

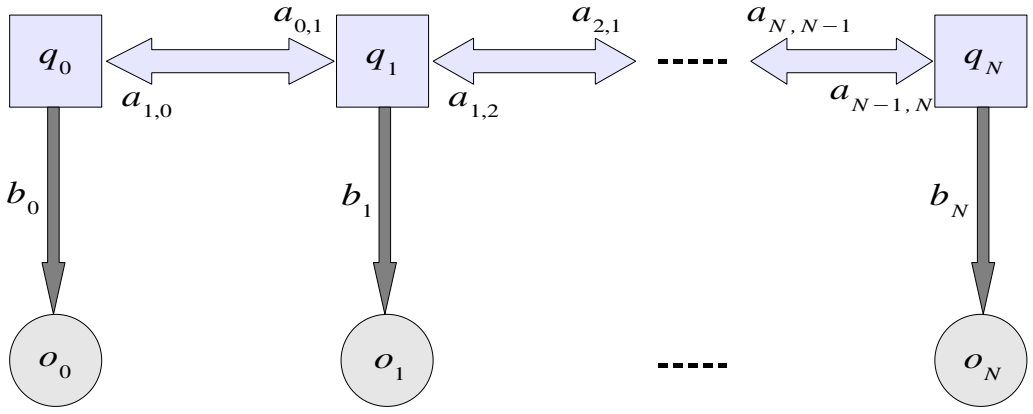


Figure 3.4: Scheme of a Hidden Markov Model (HMM). The set $\{q_i\}_{i\in\{0,\ldots,N\}}$ denotes the sequence of hidden Markov states, the set $\{o_i\}_{i\in\{0,\ldots,N\}}$ denotes the sequence of observation, and the set $\{b_i\}_{i\in\{0,\ldots,N\}}$ is the set of probability density functions (pdf). One element of this set gives the probability that one particular element of the hidden state sequence lead to to one particular element of the observation sequence .

A complete set of HMM parameters for a given model is:

$$\lambda = \{A, B, \pi\},$$

with (see Fig 3.4 on the facing page):

- $A = (a_{i,j})$ which is the transition matrix, while the elements $a_{i,j}$ denote the probability of switching from hidden state i to j,
- $B = (b_i)$ that is a vector of probability density functions in the observation space and
- $\pi = (\pi_1, \ldots, \pi_N)^T$ which is a stochastic vector, describing the initial state distribution

$$\pi_i = P(q_i = i).$$

With this at hand one has three basic problems associated with HMMs [42], [29]:

1. Determine $P(O|\lambda)$ for some sequence of observation $O = \{o(0), o(1), o(2), \ldots, o(T)\}$. This means to determine the probability of observing a certain output sequence O, given the parameter set λ.
2. Given O and any λ, one has to determine the most probable hidden state sequence (*Viterbi path*) $Q = \{q_0, q_1, q_2, \ldots, q(T)\}$. To do so, the Viterbi algorithm is used.
3. Determine $\lambda^* = \arg\max_{\lambda} P(O|\lambda)$. The Baum-Welch algorithm solves this problem (section 3.7).

In the next two sections the maximum likelihood principle and the expectation-maximization (EM) algorithm will be introduced. Afterwards we present the Baum-Welch algorithm which is the EM algorithm for HMMs.

3.4 Maximum Likelihood Principle

Consider a density function $P(O|\lambda)$ that is governed by the set of parameters λ. Especially when taking into account that P might be a set of Gaussian, λ would comprehend the means and covariances. Yet one can introduce a data set of size N

$$O = \{o_1, o_2, \ldots, o_N\}, \tag{3.25}$$

maybe generated by this distribution. Additionally one can assume that these data vectors are independent and identically distributed (i.i.d.) with distribution P. Hereon we can write down the resulting density for the samples in the following manner:

$$P(O|\lambda) = \prod_{i=1}^{N} P(o_i|\lambda) = \mathscr{L}(\lambda|O). \tag{3.26}$$

This function $\mathscr{L}(\lambda|X)$ is called the likelihood of the parameters given the data, respectively the likelihood function.

The likelihood is a function of the parameters λ where the data O is fixed. In the *"Maximum Likelihood Problem*", our goal is to find the parameter set λ that maximizes $\mathscr{L}$ [5]. This means that we have to find λ^*, in order to get the parameter estimate:

Relation 3.1 (Parameter Estimate).

$$\lambda^* = \arg\max_{\lambda} \mathscr{L}(\lambda|O). \tag{3.27}$$

Often one has to maximize the log-likelihood $\ln(\mathscr{L}(\lambda|O))$ instead of $\mathscr{L}(\lambda|O)$ for the sake of analytical and computational reasons. Analytical would mean, that we would ask for the derivatives of the likelihood function with respect to the single parameters λ and setting these derivatives equal to zero, in order to get analytical expressions for the maximum likelihood estimators. While for the problems in consideration the probability distributions of interest are complicated functions of exponential type, it is anything but easy to accomplish this analytical procedure. Numerical problems are arising here supplementary, as the likelihood function is a product of probability densities, and since these probability densities can and in fact will be very small-valued, the product function will tend to zero very fast. Building the logarithm of $\mathscr{L}(\lambda|O)$ is a solution to both problems. In the analytical case we get rid of the exponentials, and in the numerical case the problem of small numbers is vanishing, as by building the logarithm of the likelihood function the joint probability product transforms to a sum of probabilities.

The heaviness of the problem is depending on the form of $P(O|\lambda)$. In the case of Gaussian distributions where it is $\lambda = (\mu, \sigma^2)$, the problem can be solved by setting the derivative of $\ln(\mathscr{L}(\lambda|O))$ with respect to λ equal to zero, and solving directly for μ and σ^2. This will be accomplished later in the text (see section 3.10). Else wise when the problem is not so easy to solve, one has to watch for effective computational techniques to contend with the problem.

When we speak about computational methods with regard to a complex modeling problem, we have to introduce the notion of optimization. Optimization of an arbitrary problem with regard to the belonging parameter space is an important case in computational natural sciences and in general.

By this reason a short introduction to this purpose will be delivered in the next section, before we come to those special optimization algorithms that are needed for the purpose of this thesis.

3.5 Optimization

According to the general problem to find the optimal model for any physical, biological, etc. system, one is in the position to search for the optimal parameter set, that is describing the model. This will be done in the framework of inverse problems theory, which we have presented in section 3.1.

At least, when speaking in terms of computational sciences, the search for the optimal parameters for a given model is a so called optimization problem, while the computational tools, in order to accomplish optimization are the so called optimization algorithms.

Here a short introduction will be given, while important definitions of terms and nomenclature are presented in the appendix. For further information on the theme see in [46], [47] and other books, referenced in [46].

Considering a two dimensional problem space $\mathbb{X}$ and an objective function $f : \mathbb{X} \mapsto \mathbb{R}$ over this space, we could imagine a graphical representation of f, as shown in Figure 3.5. The objective function in our case could for example be given by the likelihood function.

According to this, we can distinguish between two types of optimization methods, the local optimization on the one hand, and the global optimization on the other hand. While local optimization techniques are searching the parameter space, and immediately are stopping when they reach at least a local optimum, the global optimization methods are looking for more optima within the landscape, until they maybe find the global optimum in the related parameter space.

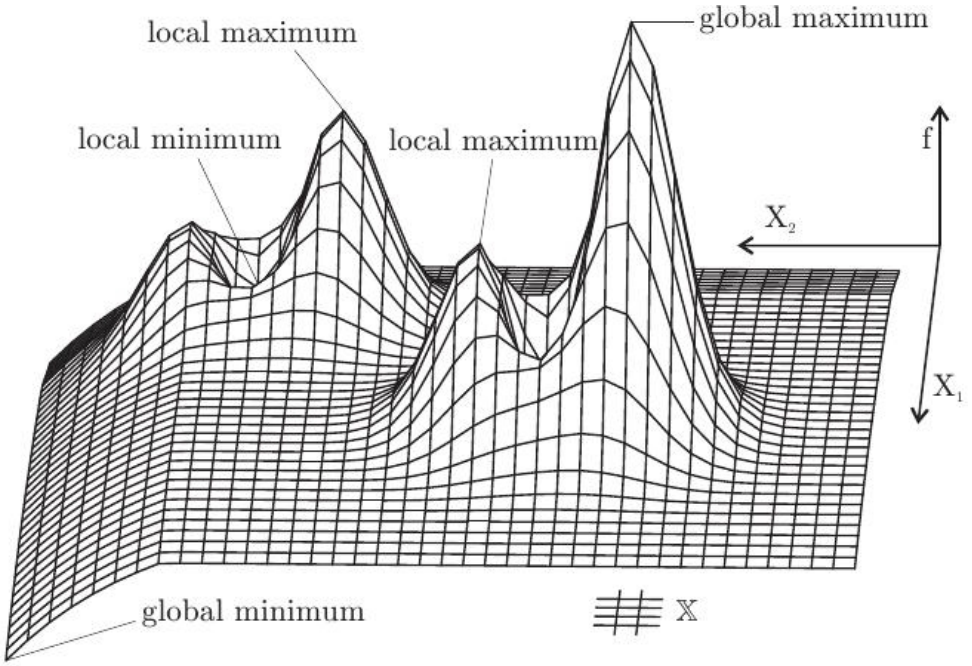


Figure 3.5: Global and local optima of a two-dimensional function f (from [46]). This function is plotted versus two elements X_1 and X_2 of the problem space $\mathbb{X}$.

An example of a local optimization (more correct maximization) method is the so called Expectation-Maximization (EM) algorithm, that will be introduced in the next section, while the Baum-Welch algorithm, which will be presented in the section after (section 3.7), is in fact the EM algorithm for HMMs.

A short introduction to global optimization methods will be given later (see section 3.13 on page 57).

3.6 The Expectation-Maximization Algorithm

The Expectation -Maximization algorithm (EM) [5] is a computational technique. This technique is a general method for finding the maximum likelihood estimate of the parameters for an underlying distribution from a given data set when the data is incomplete or has "hidden" parts. As stated in the section about inverse problems (section 3.1 on page 12), we can assume a set of observed data O and another set $X = (O,q)$, containing O and additionally a set of hidden sequence q that we do not observe. This is called the complete data.

Now we can write down a joint probability density function (pdf) for the complete data likelihood:

$$P(X|\lambda) = P(O,q|\lambda) = P(q|O,\lambda)P(O,\lambda), \tag{3.28}$$

which also leads to a new likelihood function:

$$\mathscr{L}(\lambda|X) = \mathscr{L}(\lambda|O,q) = P(X|\lambda). \tag{3.29}$$

The key object of the EM algorithm is now the expectation of the complete-data likelihood:

Relation 3.2 (EM Expectation).

$$Q(\lambda,\lambda_k) = E[\ln P(X|\lambda)|X,\lambda_k]. \tag{3.30}$$

The EM algorithm finally operates in the following way, in two steps:

Algorithm 3.1 (EM Algorithm).

- ***Expectation-Step (E-Step):*** *This step evaluates the expectation value Q based on the given parameter estimate* λ_k.
- ***Maximization-Step (M-Step):*** *This step determines the refined parameter set* λ_{k+1} *by maximizing the expectation:*

$$\lambda_{k+1} = \arg\max_{\lambda} Q(\lambda, \lambda_k). \tag{3.31}$$

$\Rightarrow$ *The maximization guarantees, that* $\mathscr{L}(\lambda_{k+1}) \geq \mathscr{L}(\lambda_k)$.

A very good introduction to the field of EM-algorithms is given in [5].

3.7 Baum-Welch-Algorithm

In this section, we derive the EM algorithm for finding the maximum-likelihood estimate of the parameters of a hidden Markov model given a set of observed feature vectors. This algorithm is called Baum-Welch algorithm. It is a special instance of the EM algorithm based on HMMs (see in [21]).

The probability of seeing the partial sequence $o_1, \ldots, o_t$ and ending up in state i at time t is given by:

$$\alpha_i(t) = P(O_1 = o_1, \ldots, O_t = o_t, Q_t = i|\lambda).$$

The variable $\alpha_i(t)$ is called the forward variable. We can efficiently define $\alpha_i(t)$ recursively as:

- $\alpha_i(1) = \pi b_i(o_1)$
- $\alpha_j(t+1) = \sum_{i=1}^{N} (\alpha_i(t) a_{i,j}) b_j(o_{t+1})$
- $P(O|\lambda) = \sum_{i=1}^{N} \alpha_i(T)$.

  This is the probability for the observation of the sequence $o_1, \ldots, o_T$ under given parameter λ. It is given by summation over all α_i at fixed time T.

The backward procedure is similar. The probability to find a partial sequence $o_{t+1}, \ldots, o_T$, given that we started at state i at time t is

$$\beta_i(t) = P(O_{t+1} = o_{t+1}, \ldots, O_T = o_T|Q_t = i, \lambda).$$

Accordingly the variable $\beta_i(t)$ is called the backward variable. Analogously we can define $\beta_i(t)$:

- $\beta_i(T) = 1$

- $\beta_i(t) = \sum_{j=1}^{N} a_{i,j} b_j(o_{t+1}) \beta_j(t+1)$
- $P(O|\lambda) = \sum_{i=1}^{N} \beta_i(1) \pi_i b_i(o_1)$

We proceed and compute the probability $\gamma_i(t)$ of being in state i at time t for the state sequence O by

$$\gamma_i(t) = P(Q_t = i|O, \lambda)$$

and express it with respect to the forward- and backward-variables. First:

$$P(Q_t = i|O, \lambda) = \frac{P(O, Q_t = i|\lambda)}{P(O|\lambda)} = \frac{P(O, Q_t = i|\lambda)}{\sum_{j=1}^{N} P(Q, Q_t = j|\lambda)}.$$

Furthermore because of the conditional independence it is:

$$\alpha_i(t)\beta_i(t) = P(o_1, \ldots, o_t, Q_t = i|\lambda) P(o_{t+1}, \ldots, o_T|Q_t = i, \lambda) = P(O, Q_t = i|\lambda)$$

and thus the definition of $\gamma_i(t)$ in terms of $\alpha_i(t)$ and $\beta_i(t)$ is:

$$\gamma_i(t) = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t)\beta_j(t)}$$

The probability of being in state i at time t and being in state j at time $t+1$ is given as

$$\xi_{i,j}(t) = \frac{P(Q_t = i, Q_{t+1} = j, O|\lambda)}{P(O|\lambda)} = \frac{\alpha_i(t) \cdot a_{i,j} b_j(o_{t+1}) \beta(t+1)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) \cdot a_{i,j} b_j(o_{t+1}) \beta(t+1)}.$$

Through the summation over all time steps t one obtains the expected total number of transitions away from state i for O :

$$\sum_{t=1}^{T} \gamma_i(t).$$

With the same assumptions one obtains for the expected number of transitions from state i to state j for O :

$$\sum_{t=1}^{T-1} \xi_{i,j}(t).$$

For the estimation of HMM we get for the relative frequency spent in state i at time 1:

$$\pi_i = \gamma_i(1).$$

The quantity a_{ij} which is an entry of the transition matrix, is the expected number of transition from state i to state j, relative to the expected total number of transition away from state i.

$$a_{i,j} = \frac{\sum_{t=1}^{T-1} \xi_{i,j}(t)}{\sum_{t=1}^{T-1} \gamma_i(t)}$$

And for discrete distributions, the quantity

$$b_i(k) = \frac{\sum_{t=1}^{T-1} \delta_{o_{t,}, v_k} \gamma_i(t)}{\sum_{t=1}^{T-1} \gamma_i(t)}$$

is the expected number of times that the output observations have been equal to $v_k$while being in state i, relative to the expected total number of times in state i.

3.8 Two Approaches on Stochastic Systems with Equivalent Results

Before we will go deeper into considerations of HMMs and the estimation processes of its parameters, the purpose of this thesis needs considerations on the physical phenomenon of diffusion, to be examined in the next section. This, as we will see can be considered by two different ways, grounding on two distinct philosophical and physical preconditions, but leading to the same results.

These two preconditions and the conditions of their equivalence are presented in and justified by the following axiom:

Axiom 3.1 (**Theoretical aproaches based on dynamics of one single system versus theoretical aproaches based on an ensemble of similar systems - Equivalence of the two assumptions jusitfied by the Quasi Ergodic Hypothesis**)**.**

Consider a stochastic system, that we can examine and for which we can set up a theoretical model. Then two models can be formulated, everyone of which based on one pre-condition:

1. *Model* 1 *based on the following assumption:*
   The dynamics of a single system in time will be described by means of a single trajectory.
2. *Model* 2 *based on the following assumption:*
   The dynamics of a single system is no more a matter of interest, moreover the dynamics of an ensemble, consisting of similar systems (trajectories), will be described.

Then the two models are equivalent in delivering the same results, denoted by

$$\textit{Model 1} \Longleftrightarrow \textit{Model 2}, \tag{3.32}$$

provided that the Quasi Ergodic Hypothesis is fulfilled. The Quasi Ergodic Hypothesis is testifying that a trajectory within the phase space of the system reaches every point in the phase space arbitrary close. The consequence is that the time average is equal to the ensemble average.

This basic separation is a central property of the so called thermodynamical systems. This means large systems, consisting of $N \geq 10^{23}$ particles. Here the question for the dynamical state (velocities $\dot{q}_k$ and momenta p_k for every single particle k) faces the question for the thermodynamical state (temperature T, volume V and pressure P).

> ”The knowledge of the thermodynamical state is by no means sufficient for the determination of the dynamical state. Studying the thermodynamical state of a homogeneous fluid of given temperature, we observe that there is an infinite number of states of molecular motion that correspond to it. With increasing time, the system exists successively in all these states that correspond to the given thermodynamical state. From this point of view we may say that a thermodynamical state is the ensemble of

> all the dynamical states through which, as a result of the molecular motion, the system is rapidly passing", as pointed out by Enrico Fermi, [20, p. 3].

In our case the dynamical states are given by the single trajectories while the ensemble state (thermodynamical state) can be represented by a probability distribution function, leading to the question how this distribution will evolve over time.

The consequence of this notion by Fermi is that the single trajectories alone on the one hand, and the ensemble state alone on the other hand will not suffice for a proper understanding of the system. By this fact, we will need both pictures. Thus in the next section a general introduction to Diffusion will be given. Afterwards as a physical motivation for the Stochastic Differential Equations (subsection 3.9.3), representing the dynamical-state approach (model 1), we introduce the Langevin Equation (subsection 3.9.1). This is denoting the dynamics of a single trajectory, driven by noise. The ensemble-state approach (model 2) will be given by means of the Fokker-Planck Equation (subsection 3.9.4), denoting the evolution of a probability density function, representing an ensemble of trajectories.

3.9 Diffusion

According to [4, p. 5] one can say that "diffusion is the random migration of molecules or small particles arising from motion due to thermal energy".

Formally one can define Diffusion by differential equations of the following type:

Definition 3.3 (Diffusion).

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2}, \tag{3.33}$$

with:

- $f(x,t)$ denoting a function in space x and time t for an ensemble of several particles of the same nature and
- D denoting a constant, the so called diffusion constant, which is depending on the nature of the particle in consideration and the environment in which it is suspended.

According to this, diffusion is a collective phenomenon. That is the motion of an ensemble, consisting of equal particles, while the single particles are moving completely independent from each other in a random manner due to thermal energy. We will see later in the text that this random motion of the single particles is the so called Brownian motion.

Furthermore diffusion is a deterministic phenomenon, deterministic with regard to the fact that diffusional motion of an ensemble always leads to equilibrium, provided that the system is isolated from the surrounding environment, and will not be disturbed externally.

Consider a basin filled with water, and in the middle of this basin, within the fluid, let there be distributed particles of the same kind in a comparatively small region (see Figure 3.6 (a)). After a finite time interval by monitoring our basin again, we observe that the area in which the particles are distributed, is augmented. Some times later it is even much more enhanced and so on, until the particles are distributed equally over the whole basin.

Mathematically expressed: When the initial distribution is delta-shaped $f(x,t_0) = \delta(x-x_0)$, then after a time $t_1 > 0$, we have a Gaussian distribution function which is broadening over time (see Figure 3.6 (b))

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

with mean μ and variance σ.

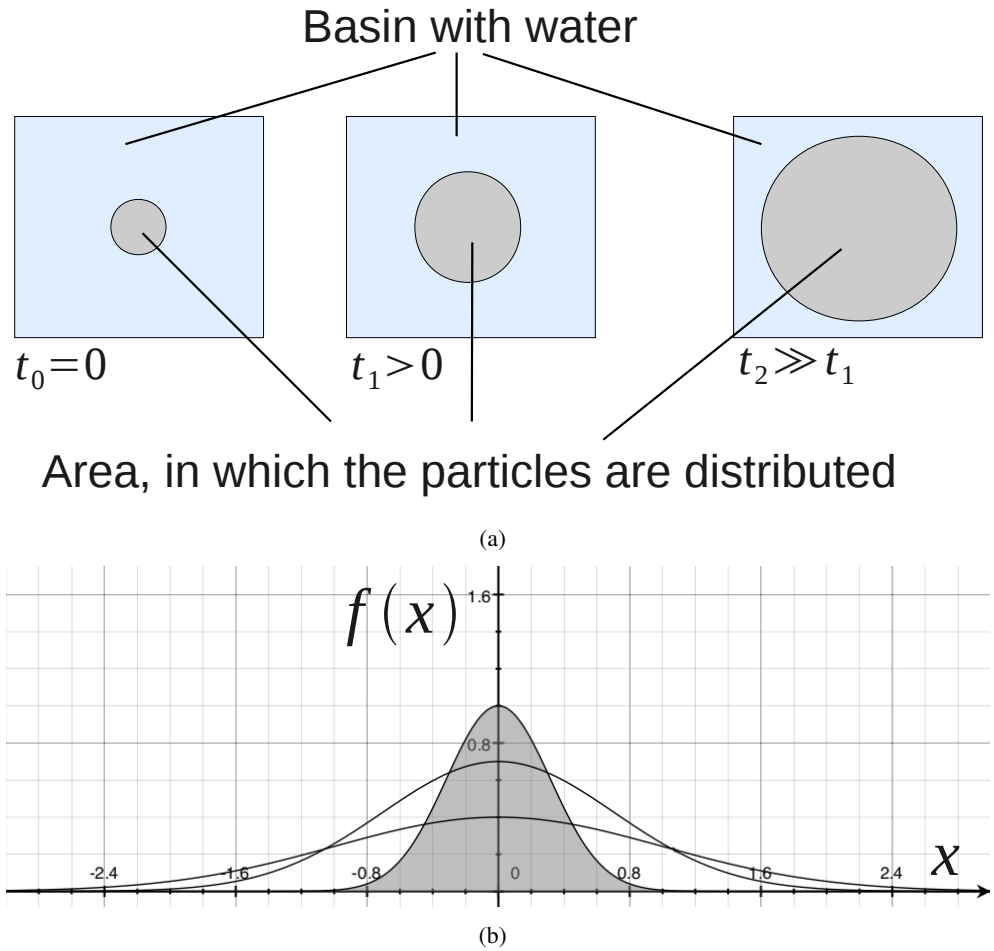


Figure 3.6: Schematic representation of the diffusion phenomenon. (a) A basin filled with water, and a particle concentration in a small area of the basin which due to diffusion propagates in time, until it is distributed equally over the whole basin. (b) Mathematically this phenomenon is represented by a Gaussian distribution function that is broadening over time.

Historically, it is not so easy to make out a definite time when the problem of diffusion entered the world of natural sciences. One reason is that the problem of diffusion is really diverse, and there are many phenomenons that can be seen and treated as diffusional movement, equal if it is taking place in a molecular environment like in cells, or if it is taking place in a more macroscopic area. The most of the literature on the topic is familiar in the opinion that it was the date, as the English botanist Robert Brown made his famous research on pollen grains, suspended in water. But it is more so that Brown rediscovered or described a phenomenon that was already known or described in the history of human science. Going back further this science history, namely to the first century A.D. when the roman philosopher Lucretius lived and wrote a scientific poem, called "De Rerum Natura" (On the nature of things), one is impressed by the following paragraph, describing the irregular movement of dust particles in the air that one can make out as a kind of Brownian motion:

> "Observe what happens when sunbeams are admitted into a building and shed light on its shadowy places. You will see a multitude of tiny particles, mingling in a multitude of ways. [...] Their dancing is an actual indication of underlying movements of matter that are hidden from our sight. [...] It originates with the atoms which move of themselves. Then those small compound bodies that are least removed from the impetus of the atoms, are set in motion by the impact of their invisible blows, and in turn cannon against slightly larger bodies. So the movement mounts up from the atoms, and gradually emerges to the level of our senses so that those bodies are in motion that we see in sunbeams, moved by blows that remain invisible."

Lucretius who stood in the tradition of the Greek philosopher Epicure, can be seen as an atomist who believed in the existence of atoms as the primary and indivisible elements of nature, in order to explain the whole existence and behavior in the universe.

Much later, in the year 1811, the brilliant physicist and mathematician Baron Jean Joseph Fourier, who at this time was the prefect of the department of Isère, France, received the price of the french academy of science for his tremendous work on the spreading of heat in solid bodies [28, p. 112]. In fact Fourier can be seen as the first who wrote down a so called heat conductivity equation, and today we know that thermal conductivity within an arbitrary material is based on the molecular movement in it, and on the other hand, as was first proved by Einstein and Smoluchowski, the so called Brownian motion or Diffusion is arising due to the thermal movement (see below). By this work Fourier can be seen as the one, who constituted the area of science, dealing with heat and its properties, today known as thermodynamics.

In 1807 Fourier gave "a careful justification of the assumption that the heat flow per unit of area is proportional to the gradient of temperature T, the constant of proportionality k (the internal conductivity) depending on the substance in question" [14, p. 165].

In this and other later works Fourier derived a differential equation "for the propagation of heat in the interior of a continuous solid,

$$cd\left(\frac{\partial T}{\partial t}\right) = k\nabla^2 T, \tag{3.34}$$

where d was the density, c the specific heat per unit and k the proportionality constant" [14, p.

169].

Rearranging (3.34) a bit leads to

$$\frac{\partial T}{\partial t} = D\nabla^2 T, \tag{3.35}$$

with $D = \frac{k}{cd}$.

It is easy to see that equations (3.33) and (3.35) are equivalent. For more information about life and work of Fourier the reader is referred to [14].

Then in the year 1827, Robert Brown was studying pollen particles floating in water under the microscope, and by this discovered that some tiny particles within these pollen grains performed some kind of irregular movements, today known as Brownian motion. Brown repeated the experiment with particles of dust and wrote down his results in a paper, submitted later in the year 1828 [44].

Although mathematical description on this phenomenon was given first by Thorvald N. Thiele in 1880, in a paper on the method of least squares, followed independently by Louis Bachelier in 1900 in his PhD thesis "The theory of speculation" in which he presented a stochastic analysis of the stock and option markets, in the end Einstein in 1905 [10] and Smoluchowski in 1906 [37], by independently from each other investigating the problem of the Brownian motion, gave a thorough mathematical description and a physical justification for the phenomenon. Einstein for example wrote at the beginning of his paper:

> ”In this paper it will be shown that according to the molecular-kinetic theory of heat, bodies of microscopically-visible size suspended in a liquid will perform movements of such magnitude that they can be easily observed in a microscope, on account of the molecular motions of heat. It is possible that the movements to be discussed here are identical with the so-called ’Brownian molecular motion’; however, the information available to me regarding the latter is so lacking in precision, that I can form no judgment in the matter” [10], [9, p. 1].

In the same paper he derives a diffusion equation which is the same as (3.33). With these important works Einstein and Smoluchowski proved indirectly the existence of atoms and molecules, which was so extremely denied by many brilliant scientists of the time, like Ernst Mach, who had a famous controversy on the existence of atoms with Ludwig Boltzmann (from 1897 to 1905, see in [32, chapter 3]).

One year later, in the year 1906, the french physicist Paul Langevin submitted a paper in which he for the first time investigated the problem of diffusion of particles from a microscopic point of view. This means that he asked for an equation of motion which describes the movement of single particles, rather than to ask for differential equations like (3.33) which is an equation for an ensemble of particles.

In the next two subsections this Langevin equation will be presented that is nothing else than a special type of stochastic differential equation (SDE), in order to have a physical motivation for the more general topic of SDEs which then follows in the third subsection (subsection 3.9.3).

3.9.1 Langevin Equation

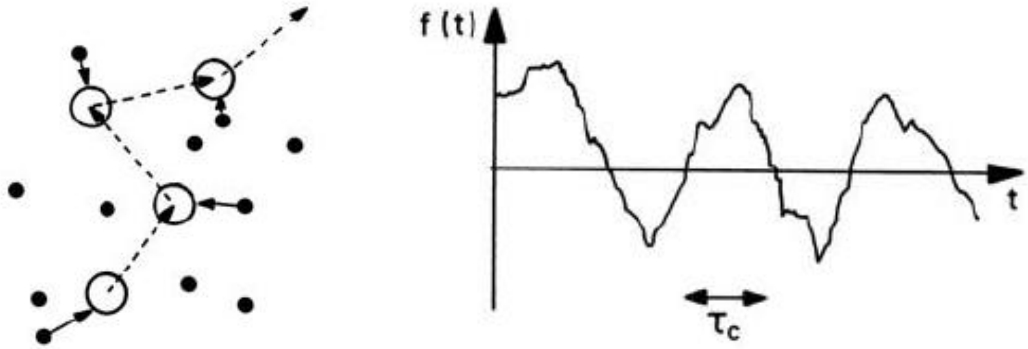
As pointed out in the last section, it was Paul Langevin the french physicist who in the year 1906 wrote down a differential equation which describes the behavior of a single Brownian particle, suspended in water. With Brownian particle it is meant that the particle is big (massive) in comparison to the molecules of water, in which it is suspended. The following motivation of the Langevin equation is based on [34, chapter 8].

A physical and mathematical elegant way to represent the Langevin equation in the scope of a first order system of stochastic differential equations is delivered in the appendix (see section A.6 on page 108).

Let m be the mass of the accordant particle, and let us assume that the particle has a velocity $v(t)$, depending on time t. Thus this velocity will change in time, according to the collisions with the molecules of water (see figure 3.7 (a)). Obviously these collisions are random in their occurrence and their overall effect is of dual character, namely an average retarding force on the one hand and on the other hand a stochastic force $f(t)$, fluctuating around this mean retarding force, as we see in figure 3.7 (b). Modeling of the retarding force via a friction term $f_{fric}(t) = -m\gamma v(t)$ delivers together with the stochastic force $f(t)$ the Langevin equation

$$\dot{v}(t) = -\gamma v(t) + \frac{1}{m} f(t) \tag{3.36}$$

which is a special type of a more general stochastic differential equation (SDE).



(a) (b)

Figure 3.7: (a) Brownian Movement and (b) Stochastic Force $f(t)$ plotted versus time t. (from [34, p. 418]).

The stochastic force is assumed to be a random variable with the following properties:

$$\langle f(t) \rangle = 0, \tag{3.37}$$
$$\langle f(t) f(t') \rangle = \phi(\tau), \tag{3.38}$$

with $\tau = t - t'$.

The question is now, how we could motivate the function $\phi(\tau)$. As we can see in figure 3.8 $\phi(\tau)$ is unequal zero only for $\tau < \tau_c$, while τ_c is the correlation time. This is the time period, the fluctuations of the stochastic force are correlated. Because we are interested in the movement of the Brownian particle for times t, which are significantly bigger than τ_c, we can approximate $\phi(\tau)$ by a delta function:

$$\phi(\tau) = \sigma\delta(\tau). \tag{3.39}$$

The coefficient σ is a measure for the strength of the mean square variance of the stochastic force. Since the friction also depends on the collisions, it is clear that there should be a connection between the noise intensity σ and the friction coefficient γ.

In order to find this connection, we have to solve the Langevin equation (3.36), which leads to:

$$v(t) = e^{-\gamma t} + e^{-\gamma t}\int_0^t d\tau e^{\gamma\tau}\frac{f(\tau)}{m}. \tag{3.40}$$

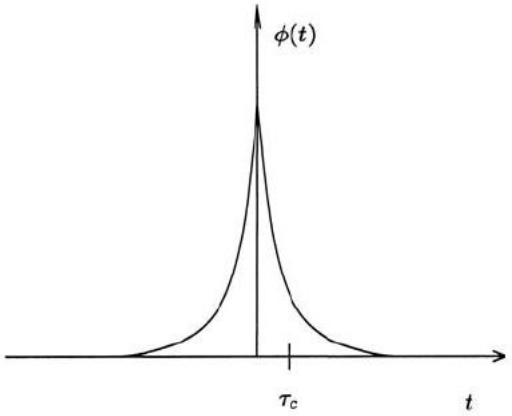


Figure 3.8: Stochastic Force Correlation $\phi(t)$ plotted versus time t. (from [34, p. 419]).

The average kinetic energy is given by

$$E_{kin} = \frac{1}{2}m\langle v(t)^2\rangle \tag{3.41}$$

while we get $\langle v(t)^2\rangle$ by using (3.39) and (3.40):

$$\begin{aligned}\langle v(t)^2\rangle &= \frac{\sigma}{2\gamma m^2}\left(1 - e^{-2\gamma t}\right) + v_0^2 e^{-2\gamma t} \\ &\overset{t\gg\gamma^{-1}}{=} \frac{\sigma}{2\gamma m^2},\end{aligned} \tag{3.42}$$

while γ^{-1} plays the role of a relaxation time. Since we are interested in times t, much bigger than this relaxation time, we can assume that the Brownian particle is in thermal balance with the environment, and thus assume that the average kinetic energy fulfills the equipartition theorem of thermodynamics

$$\frac{1}{2}m\langle v(t)^2\rangle = \frac{1}{2}k_B T, \tag{3.43}$$

where k_B is the Boltzmann constant and T the temperature of the environment, the Brownian particle is suspended in.

(3.43) with (3.42) leads finally to a special case of the *Fluctuation-Dissipation-Theorem*

$$\sigma = 2\gamma m k_B T. \tag{3.44}$$

The assertion of the Fluctuation-Dissipation-Theorem is that the magnitude of the fluctuations ($\propto \sigma$) is determined by the magnitude of the dissipation ($\propto \gamma$). For a general treatment on the fluctuation-dissipation theorem which is also called the *"Nyquist Theorem"*, see in the appendix (section A.7 on page 109).

In order to derive the mean square displacement we can first build with (3.40) the velocity correlation function

$$C_{v,v}(t-t') = \langle v(t)v(t')\rangle = \frac{\sigma}{2\gamma m^2}e^{-\gamma|t-t'|}. \tag{3.45}$$

We then get the mean square displacement by integrating (3.45) two times over time:

$$\begin{aligned} \langle x^2(t)\rangle &= \int_0^t d\tau \int_0^t d\tau' \frac{\sigma}{2\gamma m^2}e^{-\gamma|t-t'|} \\ &\approx \frac{\sigma}{\gamma^2 m^2}t \\ \Longleftrightarrow \langle x^2(t)\rangle &= 2Dt, \end{aligned} \tag{3.46}$$

with the diffusion coefficient (also denoted by the term "Einstein-Smoluchowski Relation"):

$$D = \frac{\sigma}{2\gamma^2 m^2} = \mu k_B T, \tag{3.47}$$

with the mobility

$$\mu = \frac{1}{m\gamma}. \tag{3.48}$$

The relation (3.47) is just like the relation (3.44) an instance of the fluctuation-dissipation-theorem.

Later we will be able to compute the diffusion coefficients of any trajectories by combining these two fluctuation-dissipation relations with each other, (3.47) and (3.44):

$$D = \frac{2k_B^2 T^2}{\sigma}. \tag{3.49}$$

3.9.2 Langevin Equation in an External Potential

As a generalization of the Langevin equation (3.36) we can assume that there is an external potential $U(x)$ which then leads to the Langevin equation in an external force field:

$$m\ddot{x} = -m\gamma\dot{x} - \frac{\partial U}{\partial x} + f(t). \tag{3.50}$$

We come to a special case of (3.50), namely the over-damped Langevin equation, when we assume that the friction γ is very high, thus $m\gamma\dot{x} \gg m\ddot{x}$ is fulfilled:

$$\dot{x} = -\alpha\frac{\partial U}{\partial x} + \Gamma(t), \tag{3.51}$$

with

$$\alpha = \frac{1}{m\gamma} \quad \text{and} \quad \Gamma(t) = \frac{1}{m\gamma} f(t). \tag{3.52}$$

The stochastic force $\Gamma(t)$ fulfills also (3.37) and (3.38), thus:

$$\begin{aligned} \langle \Gamma(t) \rangle &= 0, & (3.53)\\ \langle \Gamma(t)\Gamma(t') \rangle &= 2\alpha k_B T \delta(t-t'). & (3.54) \end{aligned}$$

3.9.3 Stochastic Differential Equation (SDE)

Now that we introduced a physical well known and investigated example of stochastic differential equations (SDEs), namely the Langevin equation, we should go further and define a general SDE.

Let $X(t)$ be a random variable, then we can define a general SDE as follows:

$$\dot{X}(t) = h(X,t) + g(X,t)\Gamma(t), \tag{3.55}$$

while $h(X,t)$, $g(X,t)$ are some arbitrary yet unknown functions, depending on X and t. Furthermore we have again a white noise process, denoted by $\Gamma(t)$ and fulfilling

$$\begin{aligned} \langle \Gamma(t) \rangle &= 0, \\ \langle \Gamma(t)\Gamma(t') \rangle &= 2\delta(t-t'). & (3.56) \end{aligned}$$

This stochastic process $\Gamma(t)$ is called white noise- and wiener-process or Brownian motion respectively. That is why in the following we define the Wiener process:

$$W(t) = \int_0^t dt' \Gamma(t). \tag{3.57}$$

With this we can rewrite (3.55) to the Itó SDE

$$dX(t) = h(X,t)dt + g(X,t)dW(t). \tag{3.58}$$

Then one can say that (3.58) has a solution when the integral form of it

$$X(t) = X_0 + \int_0^t h\left[X(s),s\right] ds + \int_0^t g\left[X(s),s\right] dW \tag{3.59}$$

exists for all $t > 0$.

The term $\int_0^t g\left[X(s),s\right] dW$ is the so called Itó integral. For further information about Itó SDE and integration see [26].

3.9.4 From Stochastic Differential Equations (SDEs) to the Fokker-Planck Equation (FP)

Historically, Fokker and Planck first derived a differential equation for the distribution function $\rho(x,t)$, describing Brownian motion from an ensemble point of view.

The general Fokker-Planck equation for one variable has the form [30]:

$$\frac{\partial}{\partial t}\rho(x,t) = \left[-\frac{\partial}{\partial x}D^{(1)}(x) + \frac{\partial^2}{\partial x^2}D^{(2)}(x)\right]\rho(x,t), \tag{3.60}$$

while $D^{(1)}(x)$ is called the drift coefficient and $D^{(2)}(x) > 0$ the diffusion coefficient. The drift and diffusion coefficients may also depend on time. Equation (3.60) is an equation of motion for the distribution function $\rho(x,t)$. Mathematically, it is a linear second order partial differential equation of parabolic type, and it is also called the forward Kolmogorov equation. Roughly speaking, it is a diffusion equation with an additional first order derivative with respect to x, as a comparison with the general diffusion equation (3.33) shows.

Another way to derive the Fokker-Planck equation is to write down the "Kramers-Moyal Expansion", which is an expansion of the distribution function. Afterwards it will be obvious that the Kramers-Moyal coefficients with $n \geq 3$ will vanish for the Langevin equations with δ-correlated Gaussian-distributed Langevin-forces. By this only the drift- and diffusion-coefficients will enter in the equation of motion for the distribution function. This will be done in the appendix A.8.

Without implicit derivation, the Fokker-Planck equation for the classical Langevin equation (3.36) is

$$\frac{\partial}{\partial t}\rho(v,t) = \gamma\frac{\partial}{\partial v}\left[v\rho(v,t)\right] + \frac{k_B T}{m\gamma}\frac{\partial^2}{\partial v^2}\rho(v,t). \tag{3.61}$$

The drift and diffusion coefficients are given by

$$D^{(1)} = \gamma v \quad \text{and} \quad D^{(2)} = \frac{k_B T}{m\gamma}. \tag{3.62}$$

We can write (3.61) also in the framework of an equation of continuity

$$\frac{\partial}{\partial t}\rho(v,t) = \gamma\frac{\partial}{\partial v}j(v,t), \tag{3.63}$$

with the probability current

$$j(v,t) = -v\rho(v,t) - \frac{k_B T}{m}\frac{\partial}{\partial v}\rho(v,t). \tag{3.64}$$

There is also a Fokker-Planck equation for the over-damped Langevin equation (3.51), also called the Smoluchowski equation

$$\frac{\partial}{\partial t}\rho(x,t) = -\frac{\partial}{\partial x}j(x,t), \tag{3.65}$$

with the probability current

$$j(x,t) = -\alpha\left(k_B T\frac{\partial}{\partial x} - \frac{\partial U}{\partial x}\right)\rho(x,t), \tag{3.66}$$

and the drift and diffusion coefficients given by

$$D^{(1)} = \alpha\frac{\partial U}{\partial x} \quad \text{and} \quad D^{(2)} = -\alpha k_B T. \tag{3.67}$$

3.10 Hidden Markov Models with Stochastic Differential Equation Output (HMM-SDE)

Now that we introduced the concept of Diffusion by means of the single trajectory picture on the one hand (SDEs, subsections 3.9.1 and 3.9.3) and the ensemble picture on the other hand (delivered by the Fokker-Planck equation, subsection 3.9.4), it is time to connect this to the concept of the HMMs, given in section 3.3. Afterwards, when a proper HMM-SDE model is formulated, we will go further and apply the Baum-Welch algorithm on this model. The Baum-Welch algorithm,

already presented in section 3.7, will enable us to estimate the parameter set, forming the basis of the considered HMM-SDE model.

In this case we can stick to [16] and [17].

In order to start, we can assume a Hidden Markov Model in the way that the system is arranged in an initial state, and probabilistically jumps to other states (Markov jump process) which are hidden within the scope of any observation, while it is characterized by a diffusion process within these hidden states (metastable states). Thereby for every such state we have a system of SDEs of over-damped Langevin type, in the following manner:

$$\dot{X}(t) = -\partial_x U^{(q(t))}(x(t)) + \sigma^{(q(t))}\dot{W}(t). \tag{3.68}$$

with:

- X: Random variable,
- $q(t)$: Markov jump process with states $1...M$,
- $W(t)$: Standard Brownian motion or Wiener process,
- $\Sigma = (\sigma^{(1)}, ..., \sigma^{(M)})$ contains noise intensities.
- $\mathscr{U} = (U^{(1)}, ..., U^{(M)})$ contains interaction potentials.

In the context of this work (in the framework of the rhodopsin-transducin-system) we can expect our system to have at least three hidden states:

1. Diffusion of the transducin proteins in the solution.
2. Diffusion of the transducin proteins on the membrane.
3. Diffusion of the transducin proteins on the membrane, bound to rhodopsin proteins.

Further assumptions, we should make are:

1. Assumption:

The Potentials $U^{(q)}$ are of harmonic form:

$$U^{(q)} = \frac{1}{2}F_P^{(q)}(x - \mu_P^{(q)})^2 + U_0^{(q)}, \tag{3.69}$$

while $F_P^{(q)}$ is the potential slope, $U_0^{(q)}$ the initial value and $\mu_P^{(q)}$ the mean of the potential.

2. Assumption:

Let us apply that the jump process is fixed to one state, $q(t) = q$.

3.10.1 Propagation of the Probability Density

Now consider a statistical probability density function $\rho(x,t)$ of an ensemble of solutions for the equation (3.68), and for different realizations of the stochastic process $W(t)$. As we have seen

earlier (see subsection 3.9.4), we can present for this purpose an equivalent representation of the dynamics in terms of the Fokker-Planck equation (see equation (3.60)):

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{\partial}{\partial x}[\rho(x,t) * \frac{\partial}{\partial x}U^{(q)}(x)] + \frac{1}{2}B^{(q)}\frac{\partial^2}{\partial x^2}\rho(x,t), \tag{3.70}$$

where $B^{(q)} = (\sigma^{(q)})^2 \in \mathbb{R}^1$ denotes the variance of the white noise (for $\mathbb{R}^d$ with $d \in \mathbb{Z}$ it is a positive self-adjoint matrix).

In terms of harmonic potentials like (3.69) this partial differential equation (PDE) can be solved analytically, whenever the initial density function can be represented as a superposition of Gaussian distributions like

$$\rho(x,t) = A(t)\exp(-(x-\mu(t))\cdot\Sigma\cdot(x-\mu(t))^T), \tag{3.71}$$

where $\mu(t)$ denotes the time-dependent mean of the random variable $X(t)$.

This leads to a system of ordinary differential equations (ODE) for the time-dependent parameters $\{\mu(t),\Sigma(t),A(t)\}$:

$$\begin{aligned}\dot{\mu} &= -F_P^{(q)}(x-\mu_P^{(q)}),\\ \dot{\Sigma} &= 2F_P^{(q)}\Sigma - 2B^{(q)}\Sigma^2,\\ \dot{A} &= (F_P^{(q)} - B^{(q)}\Sigma)A.\end{aligned} \tag{3.72}$$

The explicit solution of this system of equations on the time-interval $(t, t+\tau)$ is given by

$$\begin{aligned}\mu(t+\tau) &= \mu_P^{(q)} + \exp(-F_P^{(q)}\tau)(\mu(t)-\mu_P^{(q)}),\\ \Sigma(t+\tau) &= [{F_P^{(q)}}^{-1}B^{(q)} - \exp(-2F_p^{(q)}\tau)({F_P^{(q)}}^{-1}B^{(q)} - \Sigma^{-1})]^{-1},\\ A(t+\tau) &= \frac{1}{\sqrt{\pi}}\sqrt{\Sigma(t+\tau)}.\end{aligned} \tag{3.73}$$

What we can see in (3.73) is that the time-dependent mean $\mu(t)$ converges to the potential mean $\mu_P^{(q)}$ for big times τ.

In the case here, we are interested in the probability of output $O(t_{j+1})$ in metastable state $q_{t_{j+1}}$, under the condition that the system has been in the state $\{O(t_j), q_{t_j}\}$ at time t_j. For this reason we can use (3.73) with $x(t_j) = O_{t_j}$ and $\Sigma(t_j)^{-1} = 0$ which leads to the output probability distribution

$$\rho(O_{t_{j+1}}|q_{t_j}, O_{t_j}) = A(t_{j+1})\exp[-(O_{t_{j+1}} - \mu(t_{j+1}))\Sigma(t_{j+1})(O_{t_{j+1}} - \mu(t_{j+1}))^T], \tag{3.74}$$

with the solutions, given by the following relations:

$$\begin{aligned}\mu(t+\tau) &= O_{t-1} - F_P^{(q)}(O_t - \mu_P^{(q)})\tau, \\ \Sigma(t+\tau) &= \frac{1}{2\tau}B^{(q)^{-1}}, \\ A(t+\tau) &= \frac{1}{\sqrt{\pi}}\sqrt{\Sigma(t+\tau)}. \end{aligned} \tag{3.75}$$

The equations (3.75) are simplified by the assumption that we do only want to know about the evolution of the system within a short time interval $(t, t+\tau)$. This approach is called the "Euler discretization" [16].

3.10.2 The Likelihood

The considerations in the last subsection ultimately enables us to derive a joint likelihood function (see section 3.4 on page 21) for the model, given the complete data. In order to do so, we have to introduce a joint probability distribution for the observation and hidden state sequence. Consider a parameter tuple $\lambda = \{\pi, R, \mu, \Sigma, A\}$ with the initial distribution of the Markov chain π, the transition matrix R between different states of this Markov chain and additionally the three parameters μ, Σ, A, we obtain from (3.75). Now let the observed data (O_t) be given with constant time stepping τ and the probability for the transition from the hidden state i to the hidden state j given by the $ij-th.$ entry of the transition matrix $\mathscr{T} = \exp(\tau R)$. With this new transition matrix $\mathscr{T}$ and the equations (3.75) we can replace the parameters Σ, A by the parameters F_P, B, and subsequently get a new parameter tuple $\lambda = \{\pi, \mathscr{T}, \mu, F_P, B\}$, and the associated joint probability distribution for the observation O and hidden state sequence q by

$$p(O,q|\lambda) = \pi(q_0)\rho(O_0|q_0)\prod_{t=0}^{T}\mathscr{T}(q_{t-1},q_t)\rho(O_t|q_t,O_{t-1}). \tag{3.76}$$

The joint likelihood function for the model λ, given the complete data, then is

$$\mathscr{L}(\lambda) = \mathscr{L}(\lambda|O,q) = p(O,q|\lambda) \tag{3.77}$$

3.10.3 Parameter Estimation

In the sections 3.4, 3.6 and 3.7 we introduced already the methods we want to use, in order to find the definite parameter tuple of a model that maximizes the likelihood function, given the complete data. Due to analytical and computational reasons it is convenient to maximize log-likelihood function in place of the likelihood function itself. Like we already pointed out in section 3.6 on page 24, the whole maximization process computationally will be accomplished in the framework of the Expectation-Maximization algorithm and the key object of this algorithm is the Expectation (3.30). The EM algorithm finally will be executed in two steps iteratively, until the log-likelihood is maximal for the definite parameter tuple. The first step evaluates the expectation value Q based

on the given initial parameter estimate λ_k and the second step then determines the refined parameter set λ_{k+1} by maximizing the expectation (see (3.31)).

But we also pointed out in section 3.4 on page 21 that there is possible to derive the optimal parameter set, provided that we have the case of Gaussian distributions. This can be accomplished by setting the derivative of $\log(\mathscr{L}(\lambda|X))$ with respect to λ equal to zero and solving directly for the several parameters. Here it will be reproduced, what is already done in [16].

To simplify the case, we will use the notation $\lambda = \{\pi, \mathscr{T}, \mu, F_P, B\} = \lambda_k$ and $\hat{\lambda} = \lambda_{k+1}$ for the old and new parameter estimate, respectively. In order to identify $\hat{\lambda}$ we have to find the zeros of the partial derivatives of Q with respect to $\hat{\pi}$, $\hat{\mathscr{T}}$, $\hat{\mu}$, $\hat{F_P}$, $\hat{B}$. Calculation and representation of these derivatives are made much easier by introducing the so-called forward-backward variables α_t, β_t, as already introduced in section 3.7 on page 25:

$$\begin{aligned} \alpha_t(i) &= P(O_0, ..., O_t, q_t = i|\lambda), \\ \beta_t(i) &= P(O_{t+1,...,O_T|q_t=i,O_t,\lambda}). \end{aligned} \tag{3.78}$$

These variables are recursively computable with numerical effort growing linearly in T, and allow to compute the derivatives in compact form. Given these, we find $\hat{\lambda} = \arg\max_{\lambda} Q(O|\lambda)$ to be uniquely given by

$$\begin{aligned} \hat{\pi}^{(i)} &= \frac{\alpha_1(i)\beta_1(i)}{\sum_{i=1}^{M} \alpha_1(i)\beta_1(i)}, \\ \hat{\mathscr{T}}_{ij} &= \frac{\sum_{t=0}^{T-1} \alpha_t(i)\mathscr{T}(i,j)\rho(O_t|q_t, O_{t-1}, \lambda)\beta_{t+1}(j)}{\sum_{t=0}^{T-1} \alpha_t(i)\beta_t(i)}, \end{aligned}$$

$$\begin{aligned} \hat{\mu}^{(i)} &= \frac{\sum_{t=1}^{T} \alpha_t(i)\beta_t(i)(O_t - O_{t-1} + F_P^{(i)} O_{t-1}\tau)}{F_P^{(i)}\tau \sum_{t=1}^{T} \alpha_t(i)\beta_t(i)} \\ &= \frac{X_1X_2 - X_3X_4}{X_1X_4 - X_3X_5}, \\ \hat{F_P}^{(i)} &= \frac{\sum_{t=2}^{T} \alpha_t(i)\beta_t(i)(O_t - O_{t-1})(O_{t-1} - \hat{\mu}^{(i)})}{-\tau \sum_{t=2}^{T} \alpha_t(i)\beta_t(i)(O_{t-1} - \hat{\mu}^{(i)})^2}, \\ \hat{B}^{(i)} &= \frac{\sum_{t=2}^{T} \alpha_t(i)\beta_t(i)[-O_t + O_{t-1} - F_P^{(i)}(O_{t-1} - \hat{\mu}^{(i)})\tau]^2}{\tau \sum_{t=0}^{T} \alpha_t(i)\beta_t(i)}, \end{aligned} \tag{3.79}$$

with

$$X_1 = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i)(O_t - O_{t-1}), \qquad X_2 = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i)O_{t-1}^2,$$
$$X_3 = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i)(O_t - O_{t-1})O_{t-1}^4, \quad X_4 = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i)O_{t-1},$$
$$X_5 = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i). \tag{3.80}$$

3.10.4 Case of pure Diffusion

The case of pure diffusion is a simplified model which comes out of equation (3.68), when we assume the interaction potentials $\mathscr{U}$ to be flat, that is to say that the gradients would be zero.

In this special case our model is given by a system of SDEs (over-damped Langevin) like the following

$$\dot{X}(t) = \sigma^{(q)}\dot{W}(t). \tag{3.81}$$

Continuing in the framework of last section, the belonging Fokker-Planck equation would be

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{1}{2}B^{(q)}\frac{\partial^2}{\partial x^2}\rho(x,t), \tag{3.82}$$

We can solve this equation in the same way via a superposition of Gaussian distributions like (3.71) which leads to a simplified system of ODEs for the time dependent parameters $\{\mu, \Sigma, A\}$:

$$\begin{aligned} \dot{\mu} &= 0, \\ \dot{\Sigma} &= -2B^{(q)}\Sigma^2, \\ \dot{A} &= -AB^{(q)}\Sigma. \end{aligned} \tag{3.83}$$

The explicit solution of this ODE-system on the time-interval $(t, t+\tau)$ is given by

$$\begin{aligned} \mu(t+\tau) &= Constant, \\ \Sigma(t+\tau) &= \Sigma(0) + \frac{1}{2\tau}B^{(q)^{-1}}, \\ A(t+\tau) &= A(0)\exp(B^{(q)}\Sigma\tau). \end{aligned} \tag{3.84}$$

Finally we have for this case parameter-tuples $\lambda = \{\pi, \mathscr{T}, B\} = \lambda_k$ and $\hat{\lambda} = \lambda_{k+1}$ for the old and new parameter iterate, respectively.

Given these, we proceed the way we did in the last section and find $\hat{\lambda} = \arg\max_{\lambda} Q(O|\lambda)$ to be uniquely given by

$$\begin{aligned}
\hat{\pi}^{(i)} &= \frac{\alpha_1(i)\beta_1(i)}{\sum_{i=1}^{M}\alpha_1(i)\beta_1(i)}, \\
\hat{\mathscr{T}}_{ij} &= \frac{\sum_{t=0}^{T-1}\alpha_t(i)\mathscr{T}(i,j)\rho(O_t|q_t,O_{t-1},\lambda)\beta_{t+1}(j)}{\sum_{t=0}^{T-1}\alpha_t(i)\beta_t(i)}, \\
\hat{B}^{(i)} &= \frac{\sum_{t=2}^{T}\alpha_t(i)\beta_t(i)[-O_t+O_{t-1}]^2}{\tau\sum_{t=0}^{T}\alpha_t(i)\beta_t(i)}.
\end{aligned} \tag{3.85}$$

3.11 Hidden Markov Model - Vector Auto Regression (HMM-VAR) as Generalization of HMM-SDE

(This section is from [21]).

Since we have established in the last chapter the parameter estimation of a one-dimensional SDE, we have to investigate now the generalization of the HMM-SDE approach. This generalization is based on a d-dimensional SDE of the form

$$\dot{z} = F(z-\mu) + \Sigma\dot{W}. \tag{3.86}$$

The parameter estimation is obtained by investigation of an appropriate likelihood function. This extension is performed in [33]. A more generalized approach is based on Vector Auto regression (VAR) processes ([24]), here the interpretation of a SDE as VAR process is given and a comprehension is presented. A detailed analysis of the difficulties of parameter estimation of generalized multidimensional Langevin processes is given in [18].

The trajectory z_t with $t \in \{1,\ldots,T'\}$ is given at discrete points in time. According to the equation (3.86) and given an observation z_t, the conditional probability density of z_{t+1} is a Gaussian with the density function

$$f_\lambda(z_{t+1}|z_t) = \frac{1}{\sqrt{|2\pi R(\tau)|}}\exp\left[-\frac{1}{2}(z_{t+1}-\mu_t)^T R(\tau)^{-1}(z_{t+1}-\mu_t)\right], \tag{3.87}$$

with the mean and variance of the distribution given as

$$\mu_t := \mu + (z_t-\mu)\exp(\tau F)$$

and

$$R(\tau) := \int ds \exp(sF)\Sigma\Sigma^T\exp\left(sF^T\right).$$

A likelihood function with parameter set $\lambda = (\mu, F, \Sigma)$ can be constructed as

$$\mathscr{L}(\lambda|Z) = \prod_{t=1}^{T'-1} f_\lambda(z_{t+1}|z_t).$$

But unfortunately this likelihood function can not be solved analytically. Therefore we express equation (3.87) by using the definition of μ_t as

$$\begin{aligned} z_{t+1} &= \mathscr{N}\left\{\mu + (z_t - \mu)\exp(\tau F), R\right\} \\ &= (I - \exp(\tau F))\,\mu + z_t \exp(\tau F) + \mathscr{N}(0, R), \end{aligned}$$

where $\mathscr{N}$ is a multivariate normal distribution and I is an identity matrix. Using the definitions

$$\begin{aligned} \Phi &:= [(I - \exp(\tau F))\,\mu \;\; \exp(\tau F)] \in \mathbb{R}^{d\times(d+1)}, \\ X &:= \begin{pmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_{T'-1} \end{pmatrix} \in \mathbb{R}^{(d+1)\times(T'-1)}, \\ Y &:= \begin{pmatrix} z_2 & \cdots & z_{T'} \end{pmatrix} \in \mathbb{R}^{d\times(T'-1)} \qquad \text{and} \\ \varepsilon &:= \begin{pmatrix} \mathscr{N}(0,R) & \cdots & \mathscr{N}(0,R) \end{pmatrix} \in \mathbb{R}^{d\times(T'-1)} \end{aligned}$$

allows to write

$$Y = \Phi X + \varepsilon.$$

A transformation of the parameter set λ to $\lambda' = (\Phi, R)$ leads to the likelihood function

$$\mathscr{L}(\lambda'|Z) = \left(\frac{1}{\sqrt{2\pi R}}\right)^{T'-1} \exp\left[-\frac{1}{2} tr\left((Y - \Phi X)(Y - \Phi X)^T R^{-1}\right)\right],$$

for which maximum likelihood estimators are available:

$$\Phi = YX^T\left(XX^T\right)^{-1} \tag{3.88}$$

and

$$R = \frac{(Y - \Phi X)(Y - \Phi X)^T}{T' - 1}. \tag{3.89}$$

3.11.1 Estimator Calculation

The analytic estimators given in equations (3.88) and (3.89) are in general not used for the computation of the parameters, as the required matrix inversion can be numerically unstable. Instead a moment matrix is constructed:

$$\begin{aligned} M(Z) &= \sum_{i=1}^{T'} \begin{pmatrix} 1 \\ z_i \\ \vdots \\ z_{i+p} \end{pmatrix} \cdot \begin{pmatrix} 1 & z_i^T & \cdots & z_{i+p}^T \end{pmatrix} \\ &= \begin{pmatrix} XX^T & XY^T \\ YX^T & YY^T \end{pmatrix} \\ &=: \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}. \end{aligned}$$

The moment matrix is an important object, since it contains all statistical relevant information about the observed process. By rewriting the likelihood in terms of M this fact becomes evident:

$$\begin{aligned} \mathscr{L}(\Phi, R|Z) &= \mathscr{L}(\Phi, R|M) \\ &= \left(\frac{1}{\sqrt{2\pi R}}\right)^m \exp\left[-\frac{1}{2} tr\left\{\left(M_{22} - M_{21}\Phi^T - \Phi M_{12} + \Phi M_{22}\Phi^T\right) R^{-1}\right\}\right]. \end{aligned}$$

A Cholesky decomposition of the moment matrix M yields

$$\begin{aligned} M &= \begin{pmatrix} XX^T & XY^T \\ YX^T & YY^T \end{pmatrix} = RR^T \\ &= \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix} \cdot \begin{pmatrix} R_{11}^T & 0 \\ R_{12}^T & R_{22}^T \end{pmatrix} \\ &= \begin{pmatrix} R_{11}^T R_{11} & R_{11}^T R_{12} \\ R_{12}^T R_{11} & \left(R_{12}^T R_{12} + R_{22}^T R_{22}\right) \end{pmatrix}. \end{aligned}$$

Thus the estimators Φ and R can be expressed as

$$\Phi = \begin{pmatrix} R_{11}^T & R_{12} \end{pmatrix}^T$$

and

$$R = \frac{1}{m} \cdot R_{22}^T R_{22}.$$

3.12 Artificial Test Examples for Hidden Markov Model - Vector Auto Regression (HMM-VAR)

In order to test the HMM-VAR tools, developed and presented in section 3.11, it is advisable to generate artificial trajectory data with a predefined set of parameters and afterwards estimate these parameters by the HMM-VAR tools.

We generated computationally different trajectories, two- and three dimensional, based on stochastic differential equations driven by a Gaussian white noise process. These trajectories are diffusion processes with or without drift. This means that we created trajectories, representing free diffusion on the one hand and anomalous diffusion (diffusion with an external potential force) on the other hand.

The trajectories are accomplished just in the framework of an HMM-SDE model, presented in section 3.10 on page 37, and mathematically represented by the equation 3.68 on page 38.

In the next subsection we will show some results of the three-dimensional trajectory-generation process graphically, without estimating the parameters for these trajectories via HMM-VAR. After this we turn to the two-dimensional trajectories and the problem of the parameter estimation via the HMM-VAR tools (subsection 3.12.2).

3.12.1 Three-Dimensional Diffusion Trajectories

We generated three-dimensional diffusion data, based on the HMM-SDE framework (see section 3.10 on page 37), free diffusion as well as diffusion with drift, produced by a harmonic potential. The results are shown in the Figures 3.9 and 3.10 for the free diffusion and Figure 3.11 for the diffusion with potential.

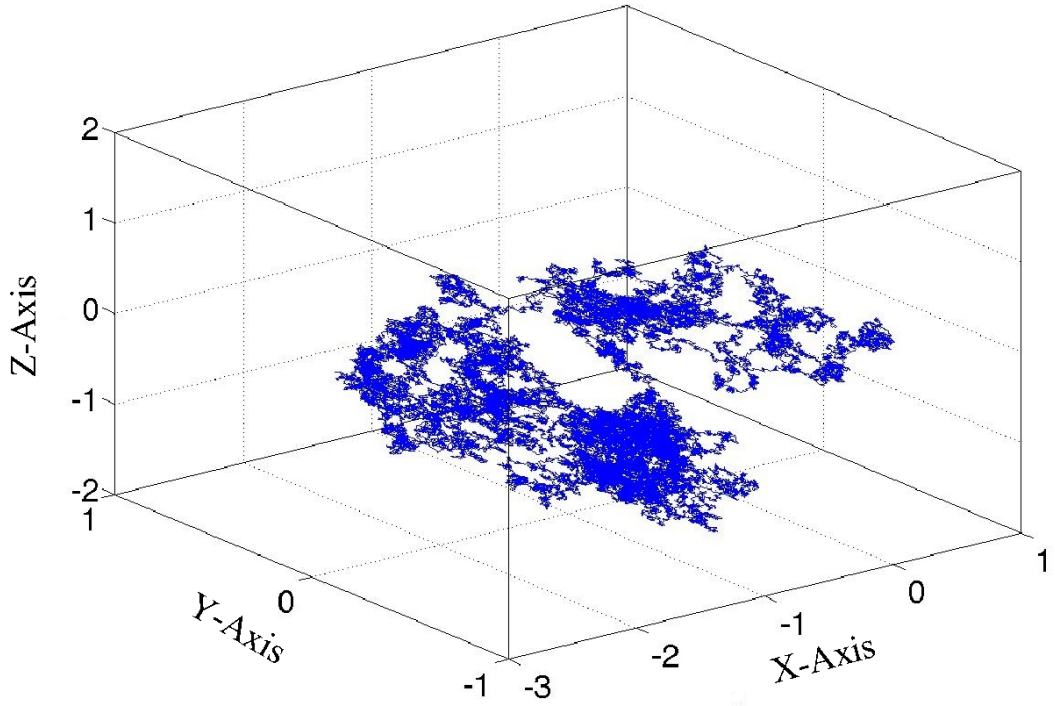


Figure 3.9: Three-dimensional free diffusion with 100.000 time steps. X, Y, Z denote the three spatial coordinates.

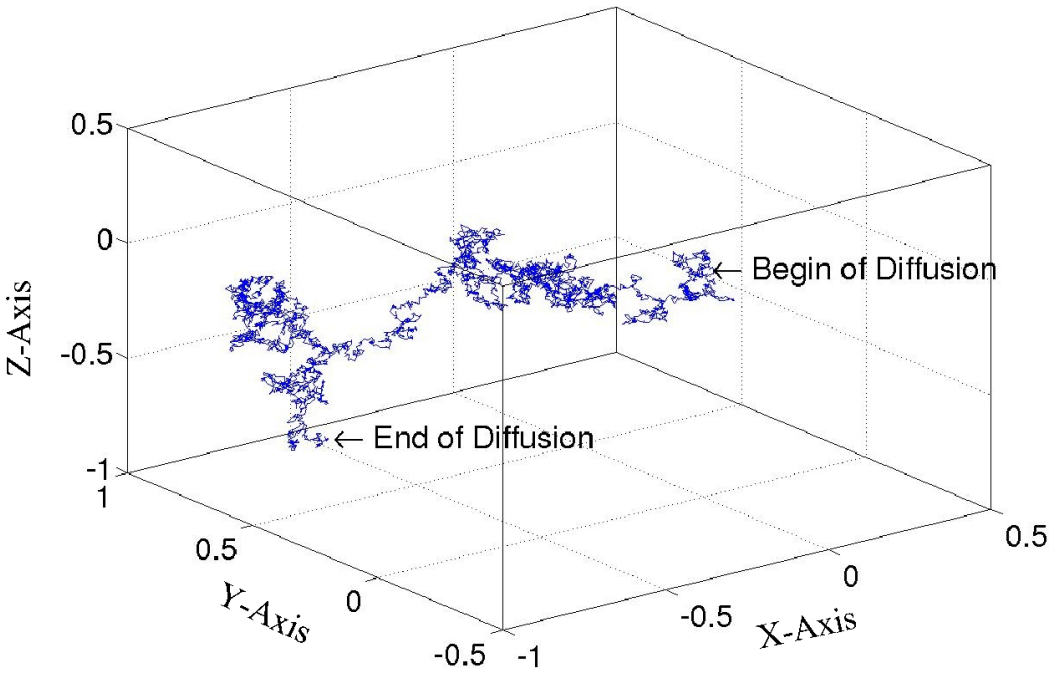


Figure 3.10: Three-dimensional free diffusion with 1000 time steps. X, Y, Z denote the three spatial coordinates.

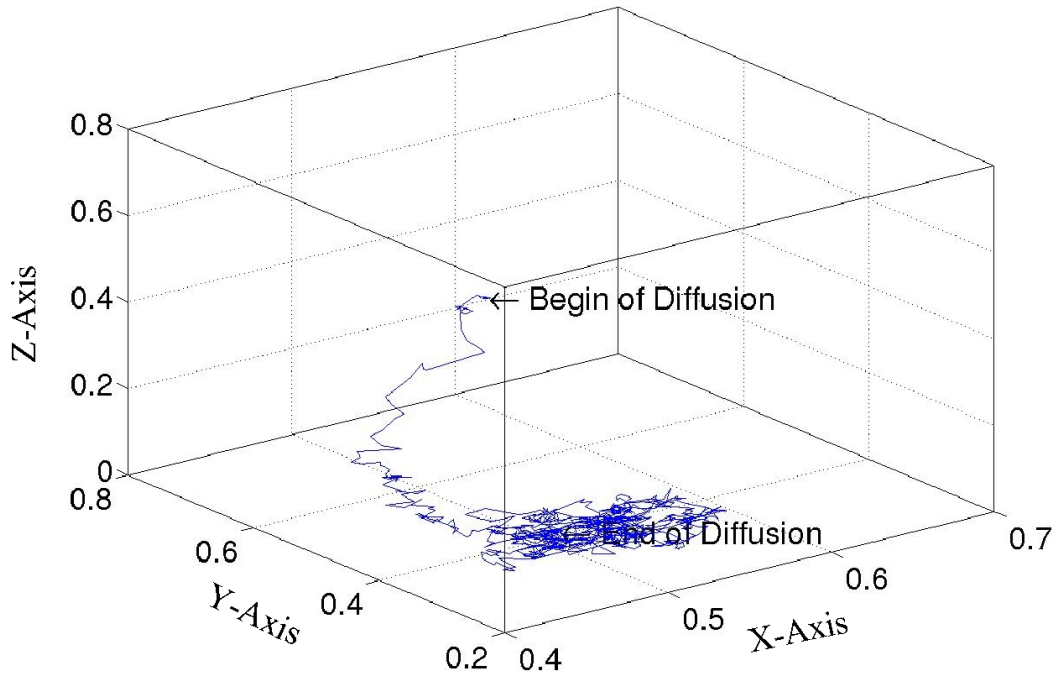
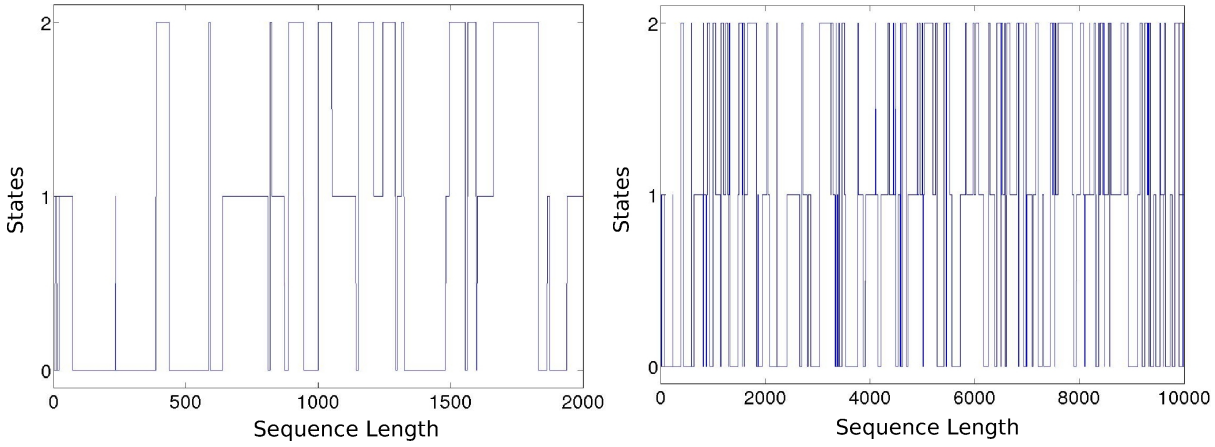


Figure 3.11: Three-dimensional diffusion with potential and 1000 time steps. X, Y, Z denote the three spatial coordinates.

As we can see in the Figures 3.9 and 3.10, the generated trajectory is representing a standard Brownian motion, while Figure 3.11 is showing a diffusional movement, which is not free, but seemingly influenced by a potential force, that is causing a drift of the process in direction of the bottom with subsequent diffusion on the vicinity of the bottom.

Just in the framework, in which these three-dimensional trajectories are generated, we will proceed and generate two-dimensional trajectories, with underlying Hidden Markov Sequences, in order to estimate afterwards these observation trajectory data. This will be done in the next subsection.

3.12.2 Two-Dimensional Free Diffusion with a Hidden Markov Model Sequence of three hidden States



(a) (b)

Figure 3.12: Random sequence of three states. Graphical representation of a random Hidden State sequence with 3 states and a sequence length of 100.000. (a) The first 2000 and (b) the first 10.000 steps.

We later will investigate experimentally observed trajectories of single molecules, in this special case single molecule tracking data of transducin proteins (see section 4.5 on page 70 and chapter 5 on page 73). These trajectory data will be two-dimensional projections of originally three-dimensional diffusional molecular movements. This fact is grounding on limitations of modern Wide-Field Fluorescence Spectroscopy in general (see section 4.3 on page 65). Thus we will apply in this section our investigations, considered the estimation of model parameter sets, to artificial examples of two dimensions. First we developed a Java based program, which generated a 100.000 steps long random sequence of three hidden states (see figure 3.12), based on the following transition matrix:

$$\mathscr{T} = \begin{pmatrix} 0.98 & 0.01 & 0.01 \\ 0.01 & 0.98 & 0.01 \\ 0.01 & 0.01 & 0.98 \end{pmatrix}, \tag{3.90}$$

and afterwards for every step of the hidden state path one diffusion-based trajectory point on the x-y-plane, based on the following parameter set:

$$\lambda^q = \{F^q, \mu^q, , \sigma^q\}, \tag{3.91}$$

with

$$
\begin{aligned}
F^1 &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & F^2 &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & F^3 &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\
\mu^1 &= \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix} & \mu^2 &= \begin{pmatrix} 0.4 \\ 0.4 \end{pmatrix} & \mu^3 &= \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \\
\sigma^1 &= \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix} & \sigma^2 &= \begin{pmatrix} 0.035 \\ 0.035 \end{pmatrix} & \sigma^3 &= \begin{pmatrix} 0.03 \\ 0.03 \end{pmatrix},
\end{aligned}
\tag{3.92}
$$

while it is of course $q \in \{1,2,3\}$, according to the respective three hidden states. Here one can see, that the potential slopes for all three states are zero, leading to the case of free diffusion, we presented in subsection 3.10.4.

For a graphical representation of a diffusion trajectory on a $2D$-plane without potential (free diffusion) see figure 3.13.

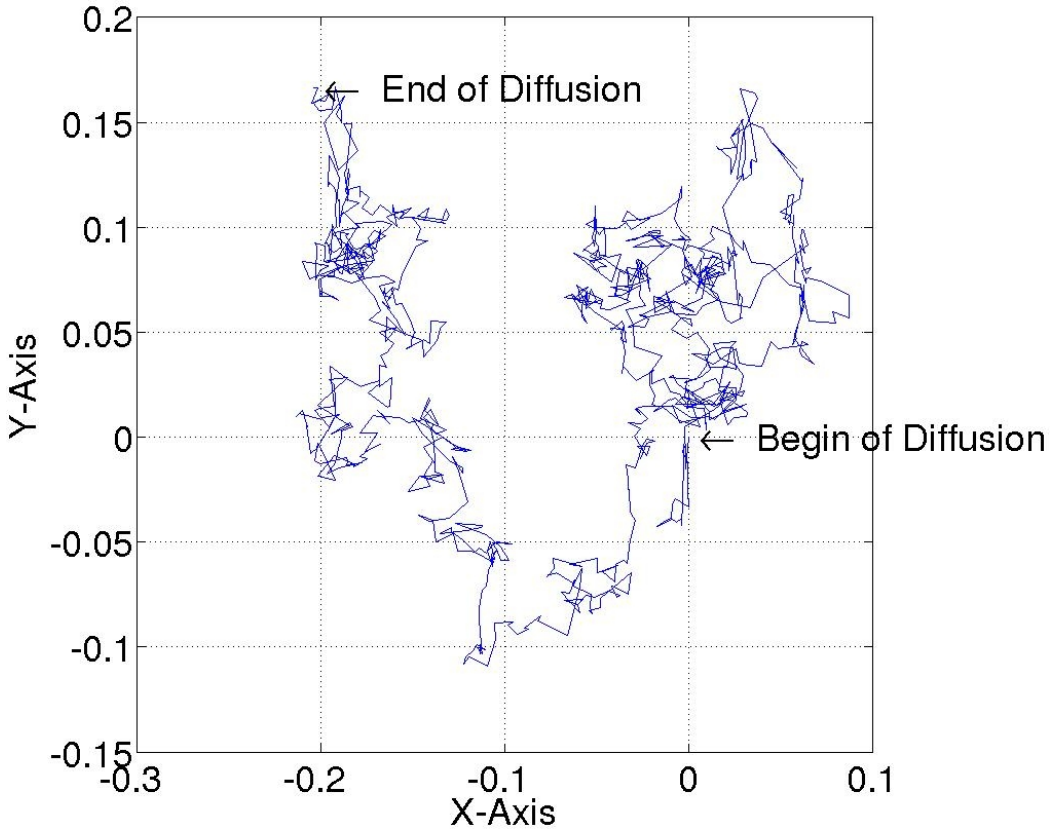


Figure 3.13: 2D free diffusion trajectory with 1000 time steps. X, Y denote the two spatial coordinates.

Afterwards, the 2D trajectory was taken as observation data, the hidden sequence as initial Viterbi path of the EM algorithm and the EM algorithm (Baum Welch) was started with a maximum number of iterations of 100000 and the precision of $1 \cdot 10^{-6}$ which is a stopping criterion for the

algorithm, meaning that it stops when the change of the new likelihood in comparison with the last one is not bigger than this number. The algorithm has to be initialized with an initial Viterbi path, in order to get started. We first decided to initialize it with a random sequence of three distinct states, generated by the same transition matrix, as the sequence which has to be estimated. The question, how the estimation result will be when initializing with a random sequence, created by another transition matrix, we will investigate later.

The procedure of estimation ended after approximately 100 iteration steps. Figure 3.14 shows the convergence of the likelihood. The estimated transition matrix that we obtained is

$$\mathscr{T}_{est} = \begin{pmatrix} 0.98 & 0.01 & 0.01 \\ 0.01 & 0.98 & 0.01 \\ 0.01 & 0.01 & 0.98 \end{pmatrix}, \tag{3.93}$$

obviously not to distinguish from the true one, given by (3.90).

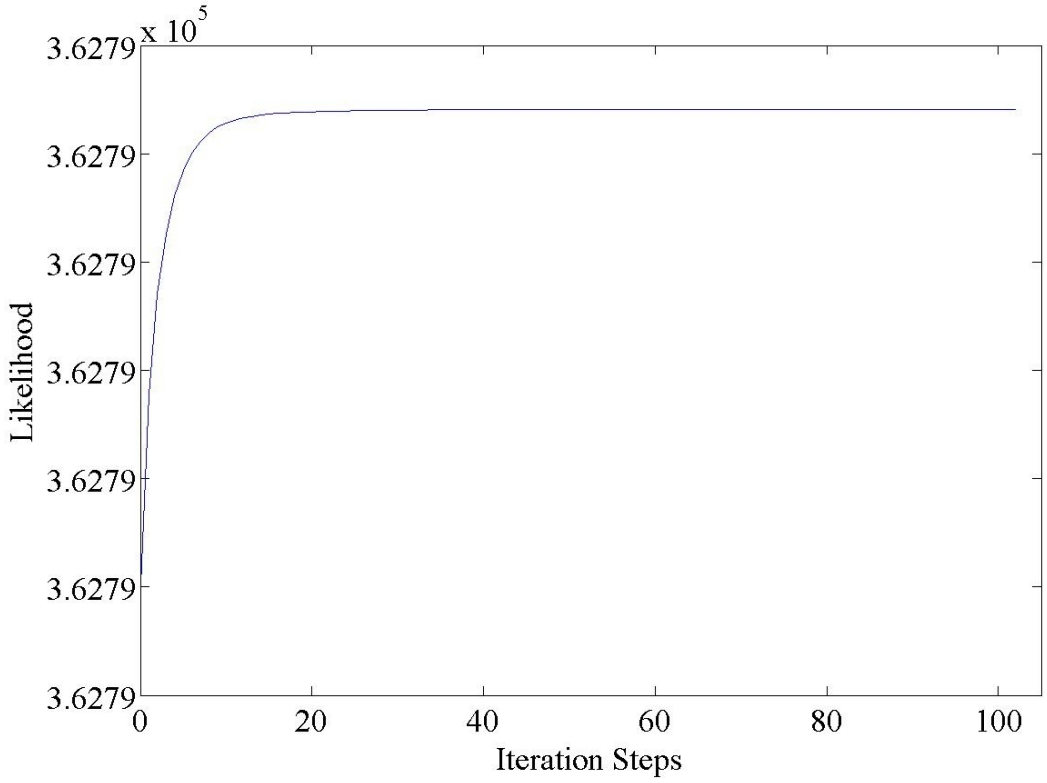


Figure 3.14: Likelihood convergence 1. Convergence of the Likelihood for the estimation of a two-dimensional trajectory with underlying HMM sequence of 100 000 steps.

We computed the distance between the two matrices, the true transition matrix (3.90) and the estimated one (3.93) which is equal to the square root of the summed squared distances between corresponding entries:

$$dist(\mathscr{T}, \mathscr{T}_{est}) = 0.002. \tag{3.94}$$

This is corresponding to a deviation of 0.2%, obviously an extraordinary good result.

Taking a much larger sequence, namely this time a sequence of length 1 000 000 and thus also a trajectory of the same length led after 130 iteration steps, for the same transition matrix, to a distance between the true and the estimated transition matrices of

$$dist(\mathscr{T}, \mathscr{T}_{est}) = 0.002, \tag{3.95}$$

which is in fact the same result as the one for a sequence with the length 100 000. Obviously a much longer sequence of hidden states does not change the result too much.

On the other hand, the question, how the result changes, when taking a much shorter sequence, namely a sequence of 1000, we also wanted to investigate and got an estimated transition matrix

$$\mathscr{T}_{est} = \begin{pmatrix} 0.99 & 4.5 \cdot 10^{-12} & 0.01 \\ 0.02 & 0.98 & 0.0 \\ 0.01 & 0.02 & 0.96 \end{pmatrix}. \tag{3.96}$$

One already is seeing by direct comparison of (3.90) with (3.96), that the estimation is not as good, as it was for longer sequences. Figure (3.15) shows the convergence of the likelihood for the estimation of the HMM-SDE data with a HMM sequence of length 1000.

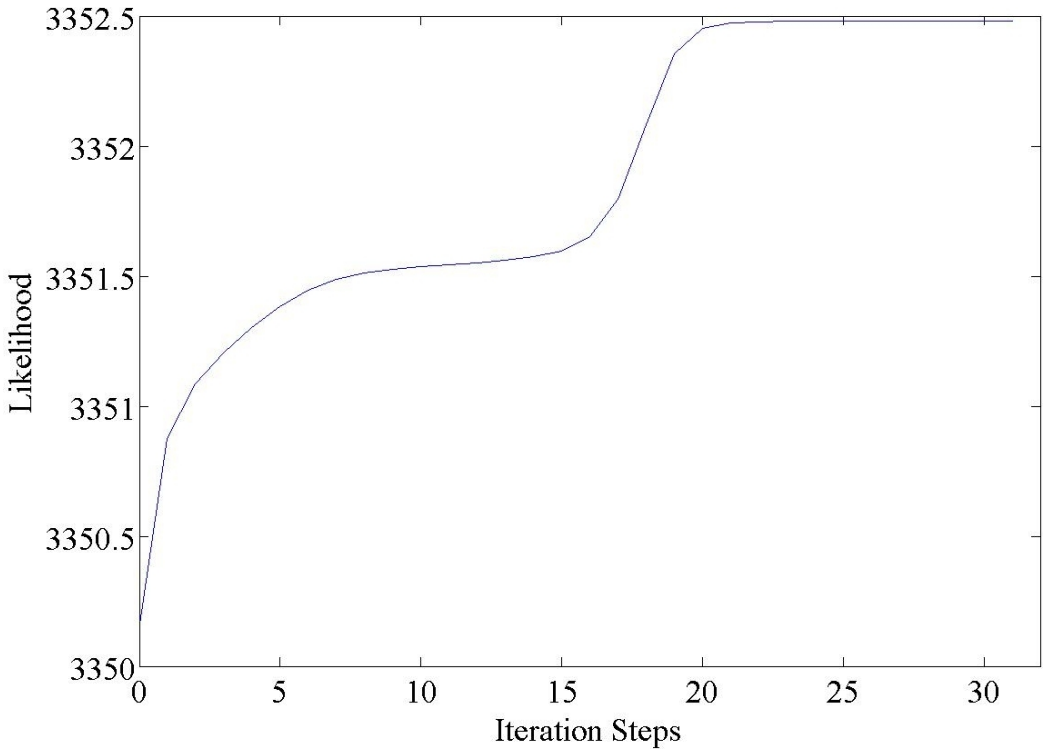


Figure 3.15: Likelihood convergence 2. Convergence of the Likelihood for the estimation of a two-dimensional trajectory with underlying HMM sequence of 1000 steps.

Next we wanted to know, how estimation results are changing, when assigning the algorithm with a random sequence of the same length, but induced by a significantly different transition matrix, for example by a matrix like the following one:

$$\mathscr{T} = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}. \tag{3.97}$$

Obviously (3.97) creates a sequence in which the states are not holding as much time, as it is the case for a sequence, induced by the matrix (3.90).

Repeated evaluation of the estimation process delivered a very interesting result. Obviously estimation is as extraordinary good, as it was for the assignment with the correct sequence. The distance between the true transition matrix (3.90) and the estimated one is approximately 0.003, corresponding to a deviation of 0.3%, although the initialization of the algorithm was done by a sequence, generated by the transition matrix (3.97), significantly deviating from the true one, (3.90).

In the following we will use this good result as a motivation, in order to go further and make some investigations on the estimation of the Diffusion constants. Since we have free Diffusion, thus there is no potential, we just can concentrate on the covariance matrices and neglect the regression

matrices for the related hidden states. Since furthermore the diagonals of the covariance matrices are equal to the variances $(\sigma^q)^2$, with q denoting the states, we can derive the belonging Diffusion coefficients, once we have σ^q, via the formula, given by equation (3.49) on page 35.

We once more generated two-dimensional Free Diffusion trajectories, with underlying random sequences of hidden states, induced by the transition matrix (3.90). The three states are denoting three states with three different noise intensity vectors, see (3.92).

With these noise intensities, the associated Diffusion coefficients can be obtained via the formula (3.49) on page 35, thus

$$D_i^{(q)} = \frac{2k_B^2 T^2}{\sigma_i^{(q)}},$$

with $q \in \{1,2,3\}$ denoting the Hidden Markov States, $i \in \{x,y\}$ the spatial coordinates, k_B the Boltzmann-constant and T the temperature.

Repeated evaluation of the estimation process delivered the correct estimation of the noise intensities and thus indirectly the correct estimation of the Diffusion coefficients.

In order to show, how the likelihood changes, when we vary the values of two particular noise intensity components, we took as an example the x-components of $\sigma^{(2)}$ and $\sigma^{(3)}$ and varied their value in an area around the correct values and evaluated for every point on the $\sigma_x^{(2)} - \sigma_x^{(3)}-$plane the associated Likelihood value. The graphical result is shown in Figure 3.16. As one can see in the Figure, the Likelihood is maximal at the $\sigma_x^{(2)} - \sigma_x^{(3)}-$point $(0.035, 0.03)$, which is the correct one.

Thus the EM-algorithm is able to detect the correct noise intensities for a particular Hidden Markov Model with Free Diffusion output, so that we are able to evaluate the Diffusion coefficients for the different states.

Until here we only investigated the case of Free Diffusion. The problem of parameter estimation for HMM-SDE with a harmonic potential is much more complicated. In this case we can not neglect the set of regression matrices, which we obtain by the HMM-VAR estimation for each Hidden Markov state, as we could assume to do for the case of Free Diffusion. Additionally these regression matrices are not independent from the covariance matrices regarding to the same Hidden Markov State. Thus determination of correct Diffusion coefficients is not really feasible.

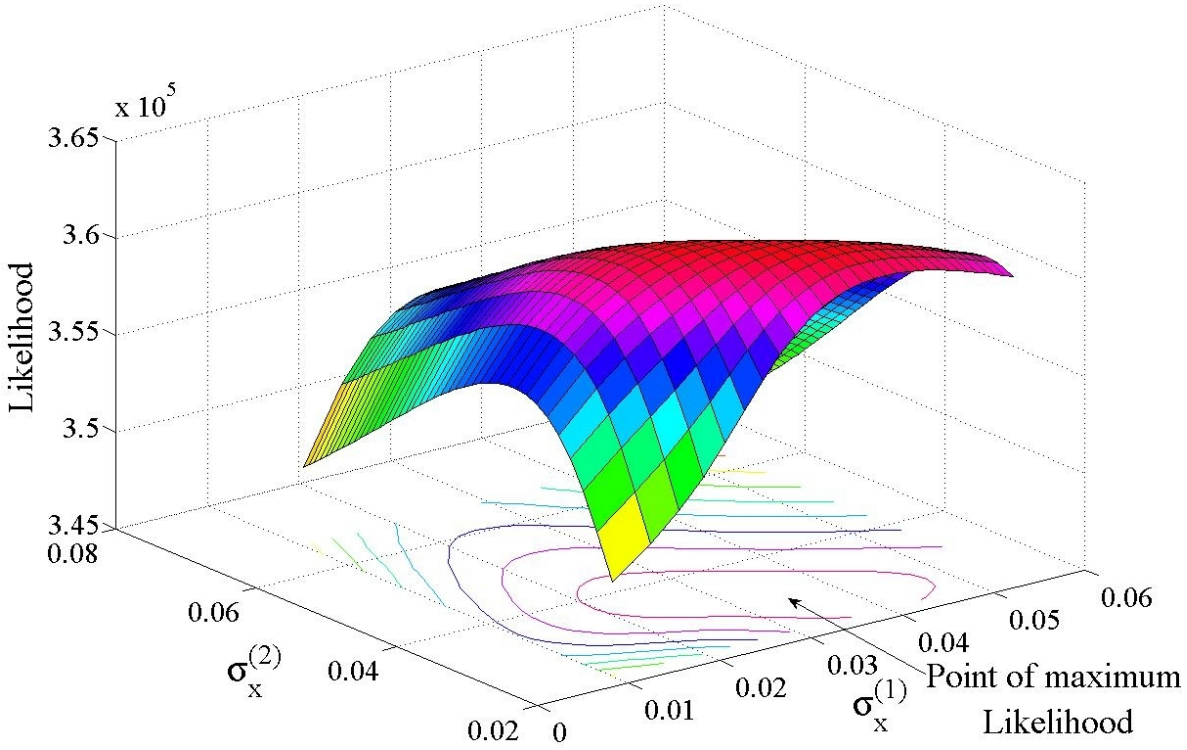


Figure 3.16: Likelihood landscape. Likelihood landscape for variations of $\sigma_x^{(2)}$ against $\sigma_x^{(3)}$, while $(2), (3)$ are denoting the Hidden Markov States 2 and 3.

We tested the estimation procedure that we already applied to Free Diffusion to the case of Diffusion with potential. First we assigned the EM-algorithm with the correct Viterbi path and afterwards with a Viterbi sequence, induced by the transition matrix (3.97). For the correct assignment the algorithm reaches quite fast the optimum and delivers the correct parameters, but for the case of initialization with significantly different sequence, the deviation between the true and the estimated transition matrix is around 50% at average and also the estimation of the other parameters fails.

3.12.3 Contemplations on Hidden Markov Models - Vector Auto Regression (HMM-VAR) Performance

As we have seen in the last section, the estimation procedure on some self generated observation trajectories with underlying hidden Markov states sequences were successful, as long as the algorithm was initialized with the correct Viterbi path. In this case it is obviously understandable, that the algorithm converged relatively fast to the correct maximum of the related likelihood.

A problem with the implementation of HMM-VAR appeared after some experiments on the parameter set λ, relative to the scales of the individual parameter values. For example taking the values of the individual noise intensities where we have seen the correct estimation of the noise intensities through HMM-VAR for the chosen noise intensities that are listed in equations (3.92 on page 50).

Variations on the scale of these noise intensities have delivered a serious drawback, namely that the algorithm crashed down, when we chose noise intensities in the magnitude of 10^{-5} and smaller, throwing a run time error, namely that the determinant of the estimated covariance matrices are zero. However this problem is not caused by HMM-VAR but is by numerical reasons. These reasons are arising with tiny numbers. For example consider the following paragraph from numerics `http://www.monash.edu.au/policy/gpnumacc.htm`:

> ”The arithmetic unit of your CPU does arithmetic extremely accurately – using 10-byte extended precision. Rounding errors arise when the result is stored (to save space) as a single precision number. For example, 4.5678 x 3.4567 = 15.78951426. This is stored in single precision as 1.578951*10^1, losing several digits. Rounding can be more noticeable in addition. For example, 10000 + 0.0001 = 10000.0001. This is stored in single precision as 1.000000*10^4 (= 10000 !)”.

So according to this, it is obvious that the tiny numbers which we chose for our noise intensities, are just rounded all to zero, causing the above mentioned problem, namely the run time error that the determinant of the estimated covariance matrices are zero.

In order to avoid this drawback, scaling can be a solution. This means that we could afterwards rescale the estimation results. We will carry this out here a bit:

So let us consider a molecular motion trajectory in the vicinity of the cell. So obviously this motion is taking place in magnitudes, that is laying in the area of nanometers up to micrometers. Furthermore let us consider these motions as Brownian, that is as diffusion processes, then experiments and theoretical research have shown, that the according diffusion coefficients are laying in the area of micrometers (see section 4.6 on page 71). These diffusion coefficients in turn give rise to noise intensities, laying in the magnitude of 10^{-35} (see equation (3.49) on page 35).

In view of the above represented problems in numeric, it is obviously impossible to estimate parameters of this magnitude via the HMM-VAR algorithm. So instead we will regard these trajectories as laying in the magnitude of 10^0, with according parameters, run the estimation procedure on the observation and afterwards rescale the trajectories and with it the according parameters.

3.13 Global Optimization Methods

We already introduced the topic of Optimization (see section 3.5 on page 23) and distinguished there between local and global optimization. Some definitions of used terms are provided in the appendix (A.4 on page 104).

The Expectation-Maximization (EM) algorithm (section 3.6 on page 24) is a representative of local optimization algorithms. This is meaning that the algorithm stops when it has reached a maximum in the related landscape of the objective function, in this case the likelihood function over the search space $\mathbb{S}$, represented by the space of parameters.

The determined maximum necessarily has not to be the global one, since by definition the parameter set λ^* has the property

$$\mathscr{L}(\lambda^*) \geq \mathscr{L}(\lambda) \text{ for all } \lambda \in \mathbb{S}, \text{neighboring} \lambda^*.$$

In opposition to local optimizing algorithms, global optimizers have the property to scan the whole search space $\mathbb{S}$, and thereby are able to find the global optimum, even if it is not necessarily guaranteed that they will do so.

There are numerous representatives of the global optimization class available. Here, we will introduce two representatives of the global optimization algorithms briefly: In the next subsection a genetic algorithm (GA), in the very next subsection simulated annealing (SA) is presented.

3.13.1 Genetic Algorithm (GA)

Genetic algorithms (GAs) form a subclass of the so called evolutionary algorithms (for definition see in the appendix A.4).

Historically, one can admit that the idea of the so called evolutionary computing was provided by I. Rechenberg, who examined the topic in his book "Evolutionsstrategie" (evolution strategy)[48]. While other researchers further developed Rechenberg's idea, it was John Holland and his research group that invented at least the genetic algorithm, treated in Holland's book "Adaptation in Natural and Artificial Systems" from 1975 [49].

One of Holland's students at this time, namely David E. Goldberg advanced the issue of GAs and wrote a book about it, "Genetic Algorithms in Search, Optimization and Machine Learning" from 1989 [50]. This is in fact the book, we can recommend to the reader, when a more detailed treatment of the topic is needed.

Without going too much into detail here, we can say that evolutionary algorithms like GAs are inspired by Darwin's evolutionary theory and central principles of this theory, such as inheritance, mutation and crossover.

Elements of the search space $\mathbb{S}$ that are subject to optimization are some individuals, characterized by a Chromosome which in turn is built up by a set of genes, while the genes finally are represented by some elements of the problem space $\mathbb{X}$, that is $x \in \mathbb{X}$.

Combining GA with HMM-VAR means that the respective HMM-VAR parameters are forming the single genes. Then one single parameter set λ composes a Chromosome of one single individual which is an element of the search space $\mathbb{S}$, as we already stated.

Taking one such individual as initial input of GA should then enable the algorithm to iteratively search and hopefully find the one individual that is holding the chromosome which represents the optimal parameter set for the model in consideration.

Taking into account the above remarks, a GA is constructed in principle as follows:[3]

Algorithm 3.2 (Genetic Algorithm).

1. ***Start:*** *Generate a random population of n individuals, for instance from an initial input individual.*
2. ***Fitness:*** *Determine the fitness of each chromosome $\lambda \in \mathbb{S}$ in the population. The fitness can be represented by an objective function $f : \mathbb{S} \mapsto \mathbb{R}$.*
3. ***New population:*** *Create a new population by repeating following steps until the new population is complete*
   a) ***Selection:*** *Select two parent chromosomes from a population according to their fitness.*
   b) ***Crossover:*** *With a certain crossover probability cross over the parents to form a new offspring (children) or make an exact copy of parents.*
   c) ***Mutation:*** *With a certain mutation probability mutate the new offspring at each position in the chromosome.*
   d) ***Acceptance:*** *Place the new offspring in a new population.*
4. ***Termination:*** *Terminate algorithm, if a certain termination criterion is fulfilled and return the best individual of the last generation, elsewise*
5. ***Loop:*** *With the current population go back to step* 2.

A further discussion of GAs, with regard to genetic operations and termination criteria in the framework of HMM-VAR is provided in [21], where also a good analysis of GA with HMM-VAR parameters as individual genes is established, and with the following results:

> ”The successful application of the genetic algorithm on top of HMM-VAR has been shown to obtain an optimal clustering, where the sole Baum-Welch has failed and got stuck in a local maximum. The obtained clustering result is here optimal in that sense, that no better likelihood has been found.
>
> By far, this does not prove the general reliability and performance of genetic algorithms, but the results have provided a first step into the successful combination of genetic algorithms and HMM-VAR.

[3] taken from http://www.obitko.com/tutorials/genetic-algorithms/ga-basic-description.php

Even though genetic algorithms seem to include a high power in the determination of global optima, they suffer from certain facts. First, no guarantee exists to really determine the global optimum. Absolute guarantee of a global minimum exists only, if the whole parameter space has been sampled. But even though genetic algorithms are just heuristics, they provide a powerful way to explore the parameter space and restrict to essential regions in that parameter space.

Especially when applied to HMM-VAR additional problems have to be taken into account:

- The parameter space of HMM-VAR parameters is large,
- numerical instabilities and problems due to improper parameters can occur during HMM-VAR calculations,
- mutation of HMM-VAR parameters is difficult,
- the likelihood landscape is far from being 'nice'."

3.13.2 Simulated Annealing (SA)

Simulated Annealing (SA) is a global optimization algorithm, originally developed by Kirkpatrick et al. [52] and has a physical background which is laying in the area of material sciences. Kirkpatrick was inspired by Metropolis et al. [51] who have developed a Monte Carlo method for "calculating the properties of any substance which may be considered as composed of interacting individual molecules".

When heating a metal and let it cool down again, the overall atomic configuration of the metal tend to a ground state, meaning that with decreasing temperature the inner energy of the system decreases, until it has reached an energy minimum (see Figure 3.17). This so called annealing process has to be slow enough, since when it is too fast, the material particles will be in an irregular configuration afterwards, and due to this the material possesses bad energy balance.

Physically the probability distribution of a configuration, characterized by a set of particle positions $\{q_k\}$, is given by the Boltzmann distribution

$$P(\{q_k\}) = \exp\left(-\frac{E(\{q_k\})}{k_B T}\right),$$

while $E(\{q_k\})$ denotes the energy of the associated configuration, T the temperature of the system and k_B the Boltzmann's constant ($k_B = 1.380\,650\,524 \cdot 10^{-23} J/K$).

While Metropolis et al. provided an exact copy of this physical process [46], Kirkpatrick et al. advanced this treatment to a global optimization method for general optimization problems, where the energy function can be replaced by an arbitrary objective function $f : \mathbb{X} \mapsto \mathbb{R}$, with $\mathbb{X}$ denoting the problem space.

The SA algorithm starts with an initial individual $p_i \in \mathbb{X}$ and one iteration randomly mutates this individual to a new one p_{i+1}, calculates the energy difference of the two individuals

$$\Delta E = f(p_{i+1}) - f(p_i)$$

and determines the acceptance probability of the new individual through

$$P(\Delta E) \quad = \quad \begin{cases} \exp\left(-\frac{\Delta E}{k_B T}\right) & ,\text{if } \Delta E > 0 \\ 1 & ,\text{otherwise.} \end{cases}$$

This relation is testifying that a new individual will be accepted, if the energy level of it is lower than the energy level for the previous individual. Otherwise, a uniformly distributed random number $r \in [0,1)$ is generated and the new individual is only accepted, if r is less than or equal to the Boltzmann probability factor.

Since for high temperatures T this factor is very close to 1, the procedure of SA allows the acceptance of many individuals, in order to allow easy escape from local optima. As the temperature decreases, fewer individuals are getting accepted and thus escape from a once reached optimum is no more possible. In the best case this optimum is a global one (see Figure 3.17).

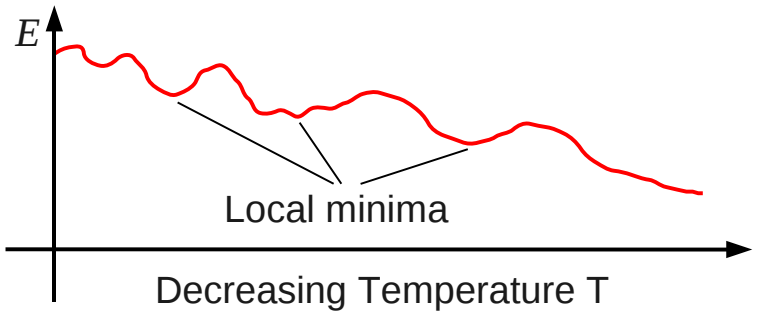


Figure 3.17: Schematic representation of the Energy function E, depending on decreasing system temperature T, and minimized via Simulated Annealing (SA) algorithm.

It has been shown that Simulated Annealing algorithms with appropriate cooling strategies will asymptotically converge to the global optimum. Nolte and Schrader [53] and van Laarhoven and Aarts [54] provide lists of the most important works showing that Simulated Annealing will converge to the global optimum if $t \to \infty$ iterations are performed, including the studies of Hajek [55].

When combining SA with HMM-VAR, the objective function would be the likelihood function $\mathscr{L}$ and the individuals p would be composed by HMM-VAR parameters, not much different than realized in the case of GA.

Both in the case of SA and in the case of GA combined with HMM-VAR, it will be the single HMM-VAR parameters that are subject to variation in order to create new individuals.

4 Fluorescence Tracking Experiments

"We run our dialog with the nature successfully from our places within the nature and nature responds only to those, who explicitly admit, to be a part of her."
(Illya Prigogine and Isabelle Stengers)[1]

In this chapter one of the main tools of today's experimentalists in the field of biophysical science will be presented, namely the method of fluorescence tracking. This will be needed in order to understand the data which we have received and we will model in the next chapter.

But before doing so we have to elaborate important terms and concepts. Thus in the next section a basic physical introduction to the phenomenon of fluorescence will be given. Afterwards we will introduce fluorescence spectroscopy as a general experimental method from which two main streams branch off:

1. Wide-Field Fluorescence Spectroscopy and
2. Confocal Fluorescence Spectroscopy.

Here the focus will be on the wide-field fluorescence spectroscopy. In the third section of this chapter the concept of single molecule tracking in the framework of wide-field fluorescence spectroscopy will be introduced. Combined with the total internal reflection fluorescence microscopy method (TIRFM, section 4.4) this will deliver the framework in which the single molecule experiments on the transducin molecules were accomplished (section 4.5).

[1] German translation: "Unseren Dialog mit der Natur führen wir erfolgreich von unserem Platz innerhalb der Natur aus und die Natur antwortet nur Jenen, die ausdrücklich zugeben, ein Teil von ihr zu sein" (Illya Prigogine, Isabelle Stengers; Dialog mit der Natur [28]).

4.1 The Physical Principle Of Fluorescence

"Luminescence" is the emission of photons by atoms or molecules [56]. If this happens after excitation by a photon of light with the energy $h \cdot \nu$, one is speaking of "photo-luminescence". Photo-luminescence is divided into "fluorescence" with relaxation times around $10^{-8}s$, occurring from excited singlet states and into "phosphorescence" with relaxation times of $10^{-3}s$ up to hours, occurring from triplet states.

An introduction to photo-luminescence in general and to fluorescence in special will be given in the following.

The properties of photo-luminescence were described first by Alexander Jablonski [56]. In the year 1935 he developed a diagram, called after him, the Jablonski-Diagram, shown in Figure 4.1.

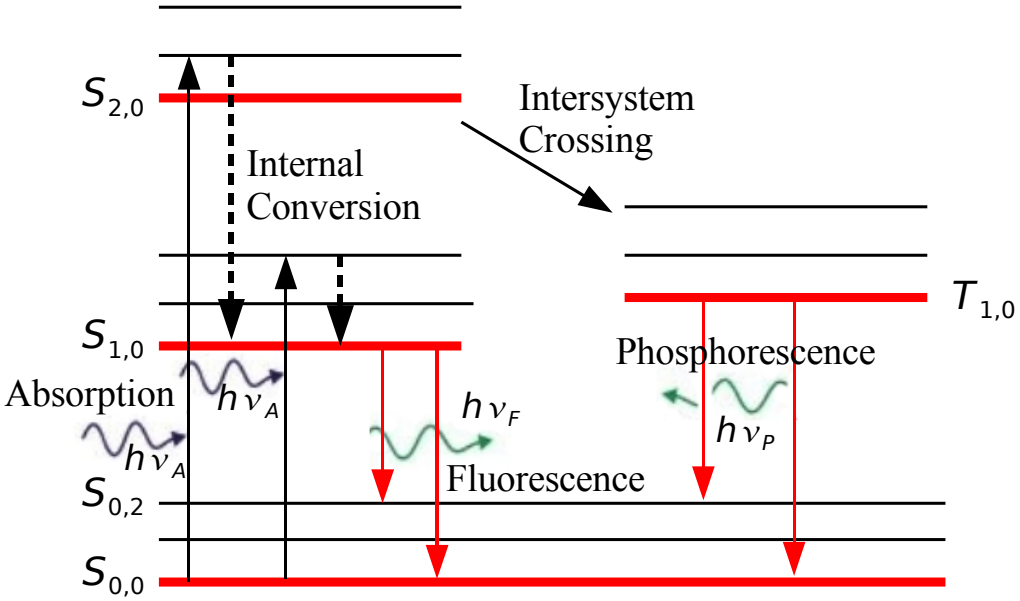


Figure 4.1: Jablonski-Diagram with schematic transitions of absorption, emission and transitions without radiation. The $S_{i,k}$ denote the singlet states, the $T_{i,k}$ the triplet states (with $i,k \in \{0,1,\ldots\}$).

A molecule with an even number of electrons can form a singlet- and a triplet-system (see Figure 4.1). A singlet state is a state with the total spin S of 0 and a multiplicity of 1, while within the triplet state the molecule has a total spin of 1 and a multiplicity of 3. The term multiplicity gives the number of possible spin orientations with respect to the $z-$axis

$$M = 2S + 1.$$

The energy levels of the molecule in the singlet state are denoted by $S_{i,k}$ and in the triplet state by $T_{i,k}$, while i denotes the excitation states of the electron and k the different vibrational states within these excitation states (see Figure 4.1).

After absorption of light with an energy $h \cdot \nu$ of an appropriate frequency ν the molecule is excited from the ground state $S_{0,0}$ to a higher oscillation state $S_{i,k}$, while the subsequent emission of light occurs from the vibrational ground states $S_{i,0}$ and $T_{i,0}$ into the ground states $S_{0,k}$ (”Kascha-Rule”). The relaxation of the molecule from the higher vibrational states to the vibrational ground states occur without emission of radiation either through ”Internal Conversion” (IC), or through ”Inter system Crossing” (ISC). IC is the relaxation of the molecule without change of spin under release of thermal radiation ($S_{i,k} \overset{IC}{\rightarrow} S_{1,0}$), while ISC is a change of the spin orientation ($S_{i,k} \overset{ISC}{\rightarrow} T_{j,k}$). The emission of light without change of spin is called fluorescence ($S_{1,0} \overset{Fluorescence}{\rightarrow} S_{0,k}$), while the state relaxation from the triplet states with change of spin is called phosphorescence ($T_{1,0} \overset{Phosphorescence}{\rightarrow} S_{0,k}$). The spin change at phosphorescence leads to an enormously longer lifetime of the triplet states compared to the singlet states which explains the long decay phase of Phosphorescence.

In general the Fluorescence spectrum is shifted to higher wavelengths compared to the absorption spectrum, a phenomenon denoted by the term ”Stokes-shift”.

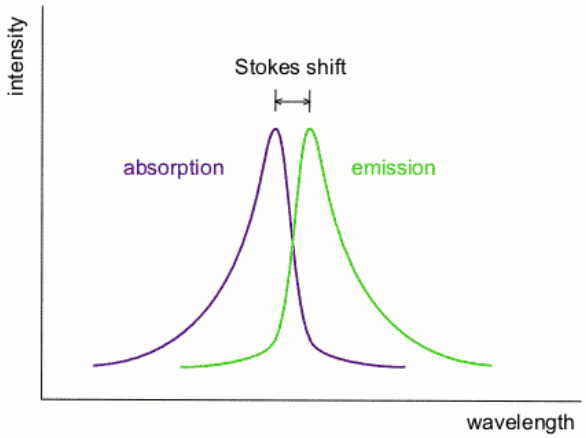


Figure 4.2: Stokes-shift of the emitted wavelength relative to the absorbed wavelength.

This is easily understandable by the following observation: imagine a molecule, suspended in an environment at room temperature or even at the physiological temperature of $37°C$ and now consider the molecule, absorbing a photon with a certain wavelength λ_a (frequency ν_a). Due to thermal fluctuations in the system the molecule is permanently vibrating, and is rarely arranged in the ground state $S_{0,0}$ and the vibrational ground state of the excited states $S_{i,0}$.

Due to this fact the molecule cannot absorb photons with energies, exactly the same as the gap energies between the ground state $S_{0,0}$ and the vibrational ground state of the excited states $S_{i,0}$. The subsequent emission of fluorescence photons is occurring after relaxation without radiation of light from vibrational excited states $S_{i,k}$ into the vibrational excited states $S_{i,0}$, and thereafter it relaxes with emission of light into the vibrational ground states $S_{0,k}$ or into the ground state itself.

For this reason the energy of the fluorescence photon E_f is lower than the energy of the absorbed photon E_a, thus the wavelength λ_f higher than the wavelength λ_a.

4.2 Fluorescence Spectroscopy

We have learned in the last section that Fluorescence is the relaxation of a molecule to the ground state under emission of a photon after excitation. The energy of the fluorescence photons is below the energy of the excitation photons [35].

Since at room temperature (or much more at the physiological temperature of $37°C$) the energy which is needed for excitation of the molecular system is much higher than the thermal energy k_BT of the environment, spontaneous transition into the excited state is highly improbable. Therefore fluorescence spectroscopy is an experimental method giving the experimentalists a tool at hand which has very high signal to noise ratios.

In the biophysical and biochemical sciences the method of fluorescence spectroscopy is of high importance. Here one is interested in any methods, in order to understand the processes that take place in biological systems, not least the processes that are going on in the cells, and within these cells the molecular constituents which are responsible for the smoothly flow of these processes.

Molecules of interest are for example proteins, molecular machines, diverse in their form and functionality, and thus making life in its thousandfold realizations possible. While proteins are built up by amino acids, these amino acids exhibit fluorescence spectra much too low in their intensities, and laying in the UV-area of the electromagnetic wave spectrum. An excitation with photons of UV would irreversibly destroy the cell.

This trouble leads to the employment of the so called fluorescence dyes, mostly small molecules with reactive chemical groups, enabling the scientists, easily "to bond these dyes to the protein of interest, just according to a kind of recipe" [35].

Not only the fact that these dyes can be easily sticked to the proteins makes them so important, but also their brilliant fluorescence properties. The problem however is the difficulty arising in order to transfer the dye-marked proteins back to the cell or into the organism to be investigated. This is a problem which led to the advent of another kind of marker molecule, the so called "Green Fluorescent Protein" (GFP).

This molecule can compete to any Laser dye with regard to the absorption cross section, the quantum yield and the photo-stability. But it has the great advantage of being genetically coded. Thus the GFP can be assembled due to gene technical strategies in almost every arbitrary cell system and subsequently can be specifically adhered to almost any protein. The enormous potential of the GFP led to the discovery of and the research on other proteins with related structure, so that to this date there exist marker proteins of different kind which can be switched on and off by light or even transform their color after radiation with short light pulses [35].

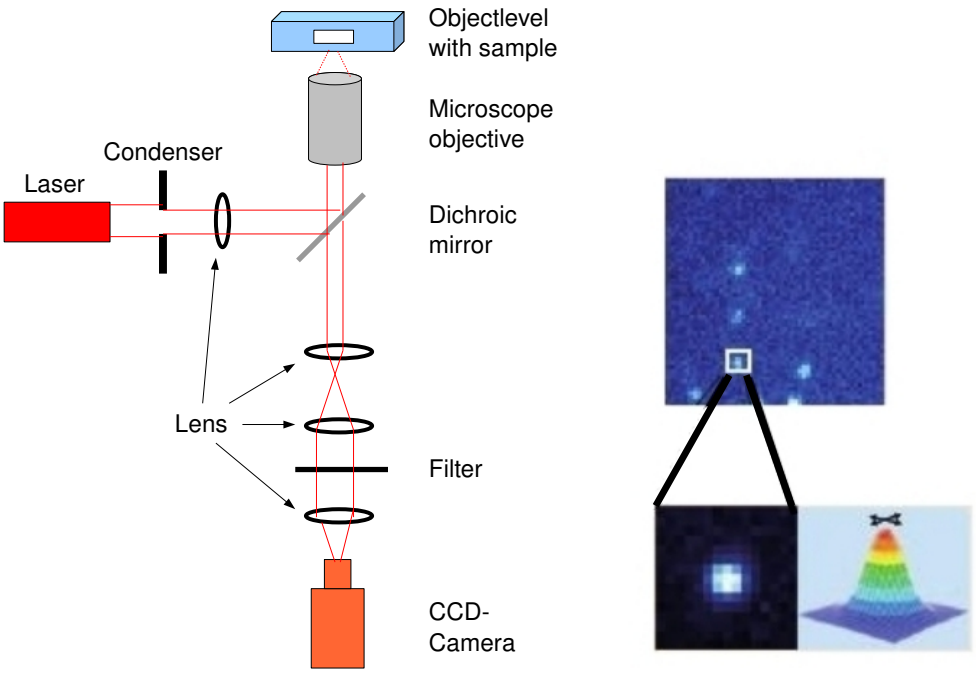
Essentially one is distinguishing between two fluorescence spectroscopy methods:

1. **Wide-Field Fluorescence Spectroscopy:**
   Planar illumination of a wide area, for example a whole cell or a cell colony. This spectroscopy method allows the investigation of different areas of a cell simultaneously, while detection of the fluorescence signals is taking place by highly sensitive CCD- or video cameras.
2. **Confocal Fluorescence Spectroscopy:**
   Point illumination is used to analyze a small measuring dot. Highly sensitive point detectors like "Avalanche Photo diodes" are imaging the signal of the measuring dot.

Since it is today technically not possible to rise the spatial and the temporal resolution of the fluorescence spectroscopy simultaneously, today's science in the field of fluorescence spectroscopy is divided in two main streams [35]. One is the so called confocal fluorescence spectroscopy deliver-

ing a temporal resolution in the investigation of biomolecular dynamics in the area of nanoseconds. We will not examine the Confocal Fluorescence Spectroscopy, since it is outside of the focus of this thesis. For more information about this method one is referred to [8]. Here we focus on the second main stream of fluorescence spectroscopy, namely the so called single molecule imaging or rather -tracking. This is a spectroscopy method, enabling the scientists to investigate the time-resolved position and movement of diverse bio molecules and other biological active particles within a cell. With this technique biological processes taking place over timescales between milliseconds and some seconds up to hours can be investigated, with possible spatial resolutions in the nanometer range ([58], [59]). Usually the single molecule tracking is done in the framework of the wide-field fluorescence spectroscopy which we will introduce in the next section.

4.3 Single Molecule Tracking via Wide-Field Fluorescence Spectroscopy



(a) Experimental Setup.

(b) Result of a measurement.

Figure 4.3: Wide-Field Fluorescence Spectroscopy (from [22]). With a wide-field fluorescence microscope for the tracking of single molecules, the light of the laser will be focused on the rear focal plane, so that the object is illuminated in the wide-field (a). The objective gathers the reflected fluorescence light and is imaging it on a highly sensitive CCD-camera. The fluorescence of the single molecules is occurring as bright dots on the CCD-camera (b1). From the detailed structure of the fluorescence image one obtains the intensity profile which can be fit by a two-dimensional Gaussian function (b2).

Wide-field fluorescence spectroscopy is described in the following according to the work of Lamb and Bäuchle [22]. As already pointed out in the last section the wide-field fluorescence spectroscopy method is suiting very well in order to track single molecules in the framework of a cell and thus mapping their movement in the cell over time which is nothing else than their trajectory within the cell. This is usually done with the help of a wide-field fluorescence microscope, where the molecules are detected as fluorescence points and are therefore resolved as individual particles, when there is no other molecule in the direct vicinity of it (see Figure 4.3).

From these single dots occurring on the CCD-camera image one can obtain the positions of the single molecules within nanometers by the center of the distribution function (see Figure 4.3 (b)), although the resolution of the optical microscope, which is limited due to the diffraction rule of Abbe, is with about 200*nm* much bigger (see for instance [58], [59]). Making many images over time (frames) and composing them to a film, one obtains the position of the particle as a function of time. Thereby one obtains the trajectory of the molecule over time.

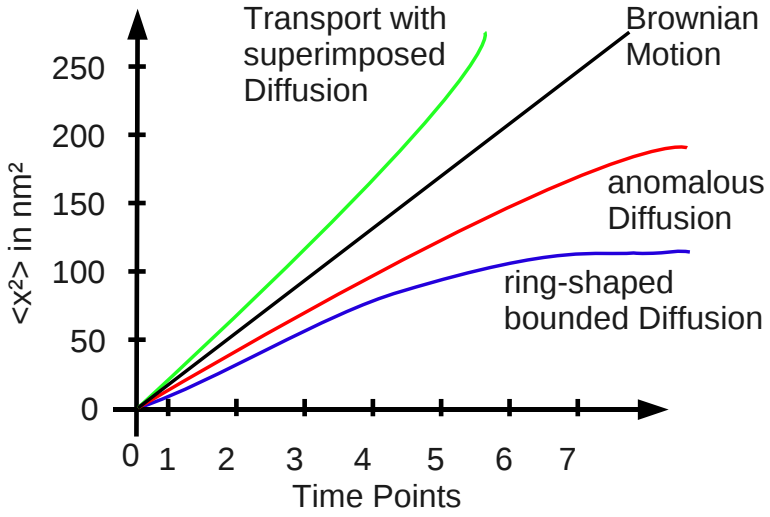


Figure 4.4: Different types of diffusion. The mean squared displacement $\langle x^2 \rangle$ is plotted versus time points.

From this trajectory one obtains not only the information about the position of the particle at every point in time, but also informations about the kind of movement which the molecule is performing. This dynamical behavior of a trajectory (for example one-dimensional movement) can be analyzed by the determination of the mean squared displacement as a function over time. Graphical representation of the mean squared displacement $\langle x^2 \rangle$ over time t shows four different diffusion types (see Figure 4.4):

1. Brownian motion is characterized by a linear relationship:

$$\langle x^2 \rangle = 2Dt,$$

with D denotes the diffusion coefficient.

2. Anomalous diffusion, which has the following relationship between the mean squared displacement $\langle x^2 \rangle$ and time t:

$$\langle x^2 \rangle = 2Dt^{\alpha},$$

where $\alpha > 1$ is a scaling factor.

3. Transport with superimposed diffusion, which is showing a parabolic behavior, characterized by the relation

$$\langle x^2 \rangle = 2Dt + \left(vt^2\right),$$

where v is denoting the velocity of the transport process.

4. Ring-shaped limited diffusion, showing an asymptotic behavior, given by the relation

$$\langle x^2 \rangle = \langle r_c^2 \rangle \cdot \left[1 - A_1 \exp\left(-\frac{2A_2 Dt}{\langle r_c^2 \rangle}\right)\right],$$

while r_c is the radius of the ring-shaped boundary and A_1, A_2 are setting deviations from the circular form of the limitation.

4.4 Total Internal Reflection Fluorescence Microscopy (TIRFM)

Single molecule detection via fluorescence spectroscopy needs a high signal-to-background ratio which is achieved through total internal reflection fluorescence microscopy (TIRFM), [39]. For the first time, TIRFM was accomplished in the year 1995, in order to detect single molecules in a solvent by room temperature [40]. Until today it is employed as promising fluorescence spectroscopy method in numerous experiments, both in vitro and in vivo [41].

TIRFM is based on the fact that after total internal reflection at the boundary surface between two media with different refraction indices n_1, n_2 and $n_1 > n_2$, a so called evanescent electromagnetic field is emerged. In order to understand this effect, we have to regard some optical properties:

The rule of Snellius says:

$$n_1 \sin \alpha_i = n_2 \sin \alpha_t$$

with

$$\begin{aligned} \alpha_i &: \text{ angle of incidence,} \\ \alpha_t &: \text{ angle of transmission.} \end{aligned}$$

It follows from the rule of Snellius the existence of a critical angle α_c:

$$\alpha_c = \arcsin\left(\frac{n_2}{n_1}\right).$$

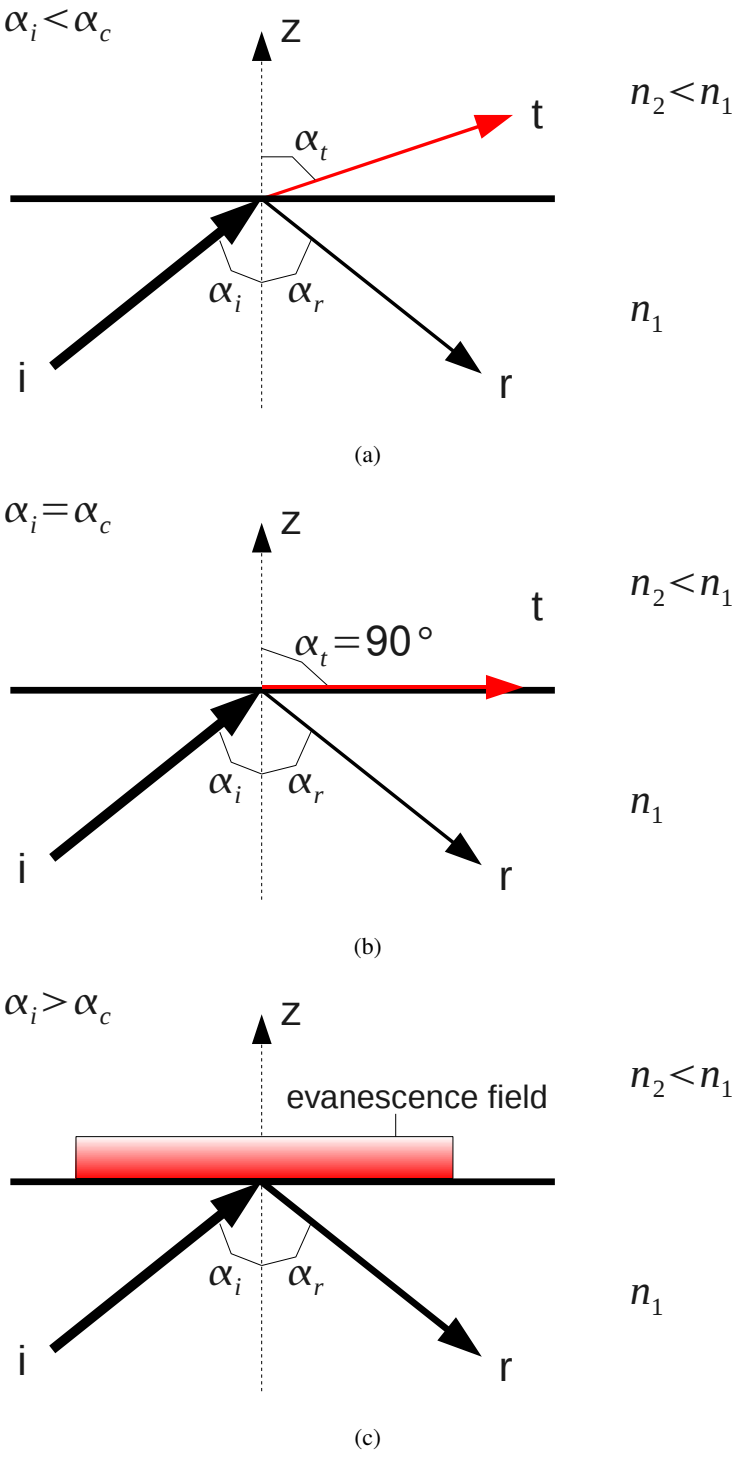


Figure 4.5: Optical behavior of incident light radiance at the boundary surface between two media with different refraction indices n_1, n_2 (i: incident radiance, r: reflected radiance, t: transmitted radiance). (a) $\alpha_i < \alpha_c$: The transmitted radiance is refracted with an angle of $\alpha_t > \alpha_i$, (b) $\alpha_i = \alpha_c$: $\alpha_t = 90°$ and the transmitted light rays are propagating along the boundary surface and (c) $\alpha_i > \alpha_c$: total internal reflection with the emergence of an evanescence field over the boundary surface.

When α_i is less than α_c, then the transmitted light rays are refracted under an angle $\alpha_t > \alpha_i$ (see Figure 4.5 (a)). When $\alpha_i = \alpha_c$, then $\alpha_t = 90°$, the transmitted light rays are propagating along the boundary surface, and a part of the incident rays is reflected under the angle of reflection α_r (see Figure 4.5 (b)). Finally, when it is $\alpha_i > \alpha_c$, one observes the above mentioned case of total internal reflection (see Figure 4.5 (c)). In this case the incident radiance is reflected completely at the boundary surface, while a tiny part of the light energy propagates along the boundary surface as an evanescent electromagnetic wave, caused through diffraction at the boundary surface.

In $z-$direction the intensity of this evanescent field decays exponentially with the depth of penetration d_p:

$$I(z) = I_0 \exp\left(-\frac{z}{d_p}\right),$$

while d_p is given by that penetration depth at which the initial intensity I_0 of the evanescent wave is declined to

$$\frac{I_0}{e}.$$

Theoretical research on and development of TIRFM was performed in the 1980s by Daniel Axelrod and colleagues (see in [43]). Two realizations of TIRFM can be differed from one another: Prism-TIRFM and Objective-TIRFM. Here we will focus on the latter.

The principle of the objective-TIRFM is shown schematically in Figure 4.6.

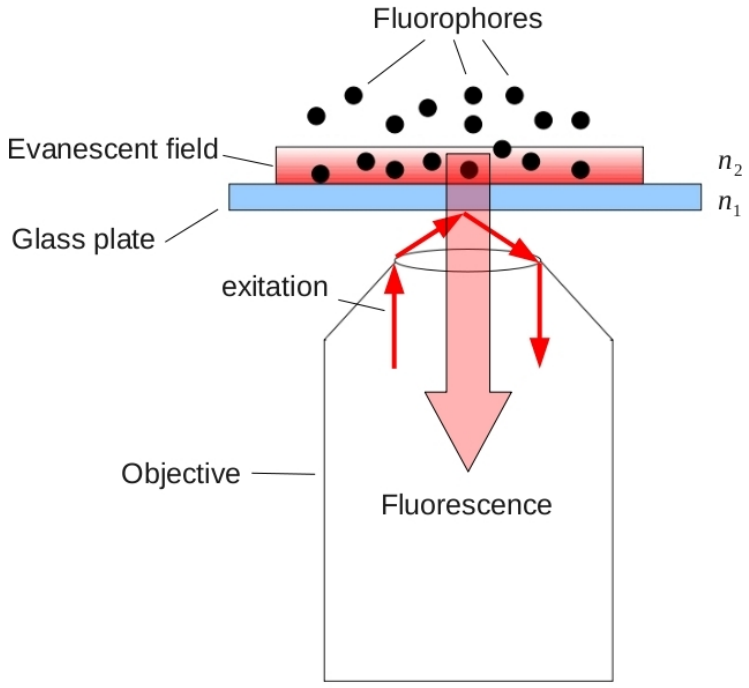


Figure 4.6: Schematic representation of objective-TIRFM. For description see text.

In the objective-TIRFM method the excitation radiance is focused through the objective on the

bottom side of the glass plate. Thereby the light rays hit the plate under an angle α_i, which is greater than the critical angle α_c, hence leading to a total internal reflection of the radiance and its recirculation into the objective.

On the upper side of the glass plate an evanescent electromagnetic wave of the same frequency, as the excitation light arises and propagates through the media above the plate, where for example fluorophores or with fluorophores labeled particles are moving.

The intensity of the evanescent wave decays with the penetration depth, and the penetration depth depends on the angle α_i under which excitation light hits the glass plate. Thus one is able to adjust the penetration depth of the evanescent field and hence adjust the size of the area which is resolved through the subsequent fluorescence light emission of the fluorophores in this area. This allows the detection of only those particles which are in the direct vicinity of the boundary surface.

The TIRFM method combined with wide-field detection through sensitive CCD-cameras delivers a good method in order to resolve single molecules at room temperature, especially in the direct proximity of the membrane. This will be presented in the next section by means of the single molecule tracking experiments on the G-Protein transducin.

4.5 Tracking Of the Transducin Proteins via Wide-Field Fluorescence Spectroscopy combined with Total Internal Reflection Fluorescence Microscopy (TIRFM)

In this chapter we described the experimental techniques that allow tracking of single molecules, and how to obtain trajectory data via these experiments. The experimental setup for the special purpose to track single transducin molecules (performed by Alexiev[2] et. al.) is based on the wide-field fluorescence spectroscopy (WFFS) method, shown in Figure 4.3 on page 65, combined with the total internal reflection fluorescence microscopy (TIRFM) method, schematically shown in Figure 4.6. The probe is illuminated via an objective in the wide-field at an angle, which is above the critical angle, leading to a total internal reflection of the light and emergence of an evanescent field at the upper boundary surface.

The probe preparation is shown in Figure 4.7 on the next page.

Bovine retinal rod disc membranes were cracked open and then placed on the bottom of a glass container filled with a solvent. The solvent contains diffusing single transducin molecules which are labeled with fluorophores. They are absorbing light of the same frequency as the evanescent wave. Afterwards emitted fluorescence light is collected by the objective. For the detection of this fluorescence light a sensitive CCD-camera is used. The time-resolved recording of the fluorescence signals of single transducin molecules lead to a two-dimensional trajectory in time. The TIRFM setup causes only the transducins in the direct proximity of the disc membranes to be resolved.

Additionally the membrane integral proteins rhodopsin can be activated through an additional laser with a different frequency. The state where the rhodopsins are not activated, will be denoted by the

[2]For contact informations look at http://users.physik.fu-berlin.de/~alexiev/

term ”Dark State”, while the activated state of the rhodopsin will be denoted by the term ”Bright State”.

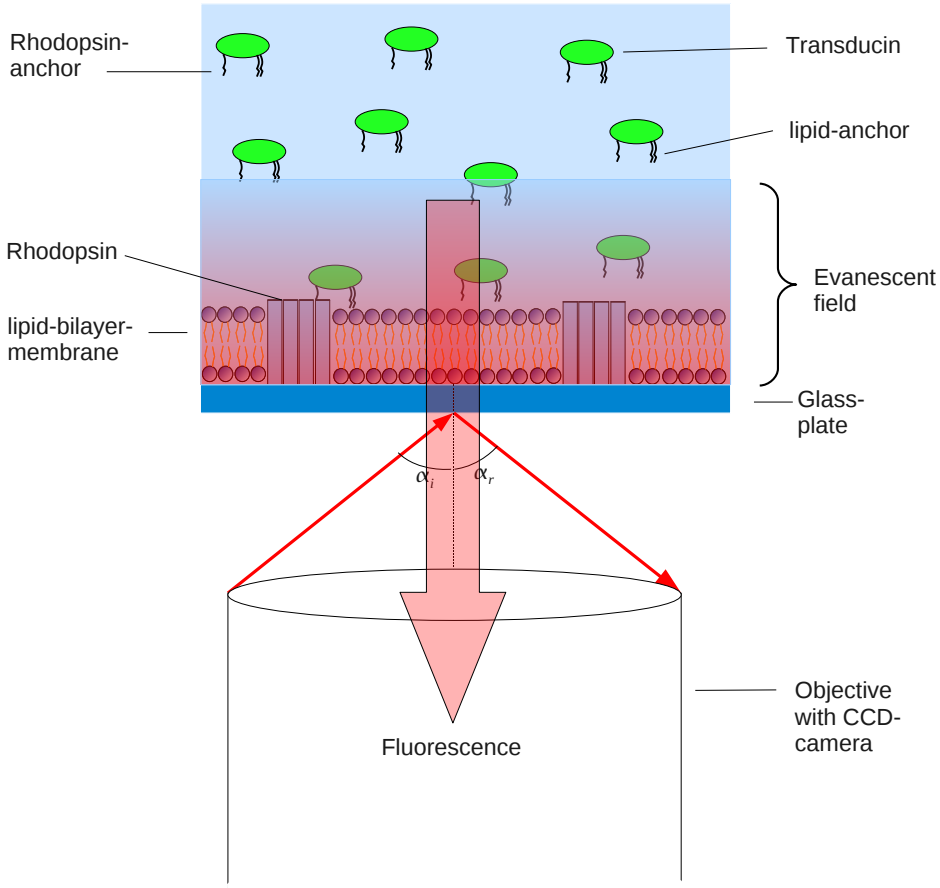


Figure 4.7: Experimental preparation of disc membranes, in order to get the movement of single transducin proteins within a solvent. Description in text.

4.6 The expected range for the Transducin Diffusion Coefficients

Already very early works on the rhodopsin-transducin system, such as the one from the year 1989 by Saxton and Owicki, have investigated the rhodopsin-transducin system (see [31]). Later works also investigated the case thoroughly (see in the appendix, section A.2 on page 101 for a presentation of selected works on the rhodopsin-transducin system).

Saxton and Owicki investigate the question how the concentration of the considered proteins affect

the reaction rates. For this reason they use in their work a wide range of diffusion coefficients, both for rhodopsin and transducin, beginning with values at infinite dilution to those values at much more dense protein concentration. For rhodopsin they give diffusion coefficients in the range of

$$D_R = 0.35\,\frac{\mu m^2}{s}, \ldots, 5.0\,\frac{\mu m^2}{s}. \tag{4.1}$$

They are pointing out that no diffusion coefficients for transducin were measured, but coefficients for other peripheral proteins (D_P) similar to transducin in the range

$$D_P = 5.0\,\frac{\mu m^2}{s}, \ldots, 6.9\,\frac{\mu m^2}{s}. \tag{4.2}$$

Saxton and Owicki, and also later scientists point out that transducin at least moves as fast as rhodopsin but moreover faster. [6] for instance calculated a lower bound for the transducin diffusion coefficient with a value of $D_T^{low} = 0.8\,\frac{\mu m^2}{s}$.

In view of this considerations we have decided to choose for the transducin molecule diffusion coefficient a relatively wide range, given by

$$D_T = 0.8\,\frac{\mu m^2}{s}, \ldots, 7.0\,\frac{\mu m^2}{s}. \tag{4.3}$$

In the following chapter we will focus on the modeling procedure based upon the experimental transducin trajectory data. We will stick to the above range in which the estimated diffusion coefficients should also lie.

If we calculate the range of noise intensities σ_T according to the determined diffusion coefficients, we obtain

$$\sigma_T = (0.5, \ldots, 7.0) \cdot 10^{-35}\,\frac{Kg^2m^2}{s^3} = (0.5, \ldots, 7.0) \cdot 10^{-35}\,\frac{J \cdot Kg}{s}, \tag{4.4}$$

where J is denoting joule, the unit of energy (heat).

5 Modeling of the Experiment

"The Begin of all sciences is the astonishment

that the things are, like they are."

(Aristotle)[1]

In the last chapter the experimental methods which were used, in order to track single transducin molecules in a solvent above the membrane were presented.

As we pointed out in section 4.5, the experiments on the movement of the transducin molecules can be separated in two different experimental setups, namely the "Dark State", denoting the situation when the rhodopsin molecules on the membrane are not activated, and the situation when the rhodopsin molecules are activated, denoted by "Active State". First we will investigate the dark state in the following, and afterwards we will come to the active state.

In the first section of this chapter we will present the related trajectories graphically. This may help us to understand what type of observation we have and how we can model it.

In the next section (section 5.2) a model ansatz will be introduced, discussed and a parameter estimation on the basis of this model will be accomplished. Discussion of the results will also be delivered here, while testing of the model will be performed in section 5.3.

Related to the transition matrices we will use the term "holding probability" when we mean the diagonl entries of the considered matrices. This holding probabilities denote the probabilities that a certain state makes a transition to itself and not to another state.

Further analysis of the data based on the separate estimation of the x- and y-components and separate estimation of the single trajectories will be delivered in the sections 5.4 and 5.5.

A theoretical development of an estimation framework based on Hidden Markov Model - Stochastic Differential Equation with differentials d_x in place of positions x will be given in section 5.6. The associated implementation in the framework of an object oriented programming language in order to estimate data for model parameters will not be delivered. This is a task for future works.

[1] German translation: "Der Beginn aller Wissenschaften ist das Erstaunen, daß die Dinge sind, wie sie sind" (Aristoteles).

5.1 Experimental Data

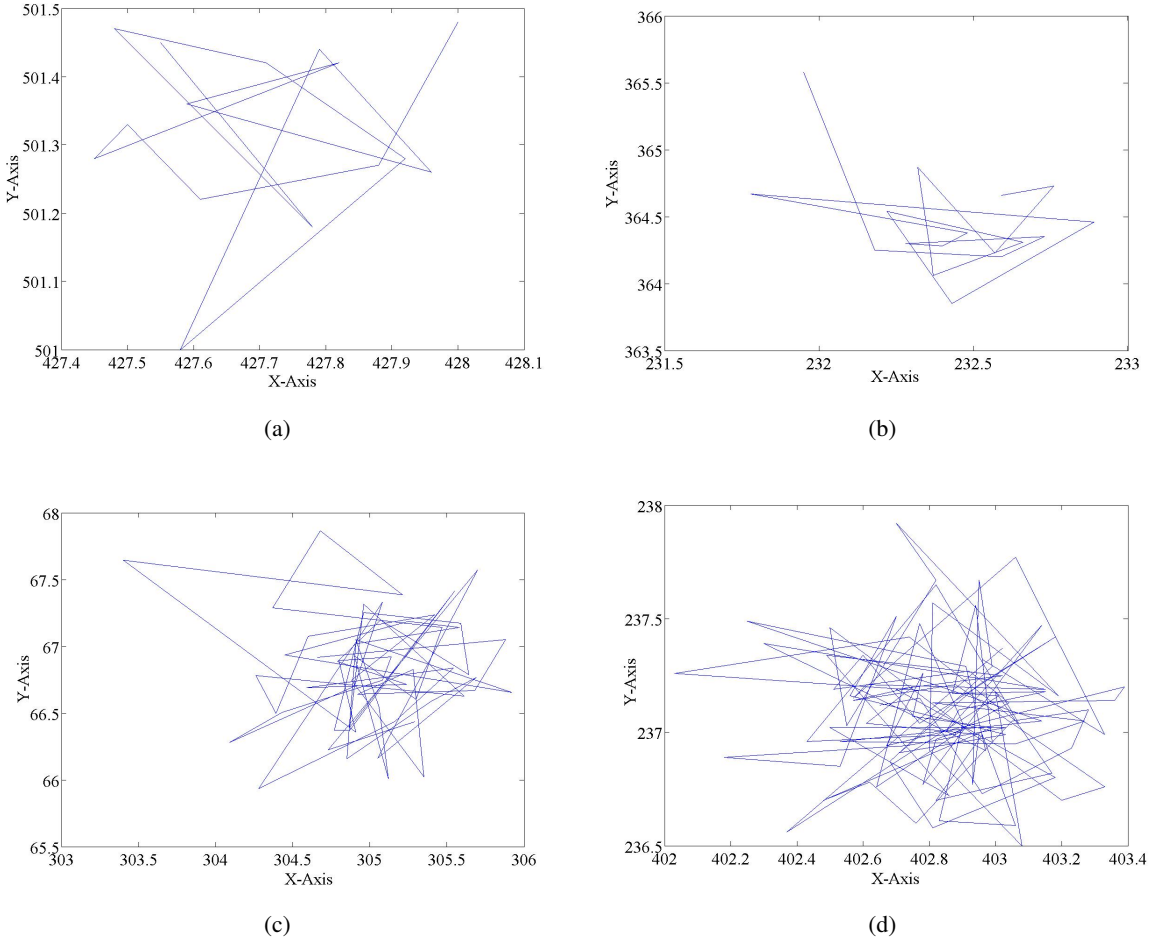


Figure 5.1: Transducin trajectories. X, Y denote the two spatial coordinates. Short trajectories with a) 15 frames, b) 18 frames and long trajectories, with c) 57 frames, d) 100 frames. Range of movement: $1.5 \times 1.5 \mu m$.

We have received data from single molecule experiments on the transducin molecules. This data which is consisting of 5721 data points, is a set of single two-dimensional trajectories. we first plotted some of this trajectories, two rather ”short” trajectories (Figures 5.1 (a) and (b)) and two ”long” trajectories (see Figures 5.1 (c) and (d)).

Afterwards, we plotted all of this single trajectories together in one plot (Figure 5.2), and four small regions within this general plot were additionally plotted (see Figure 5.3) in order to show that the small points in the general plot are the single transducin trajectories (see Figure 5.1).

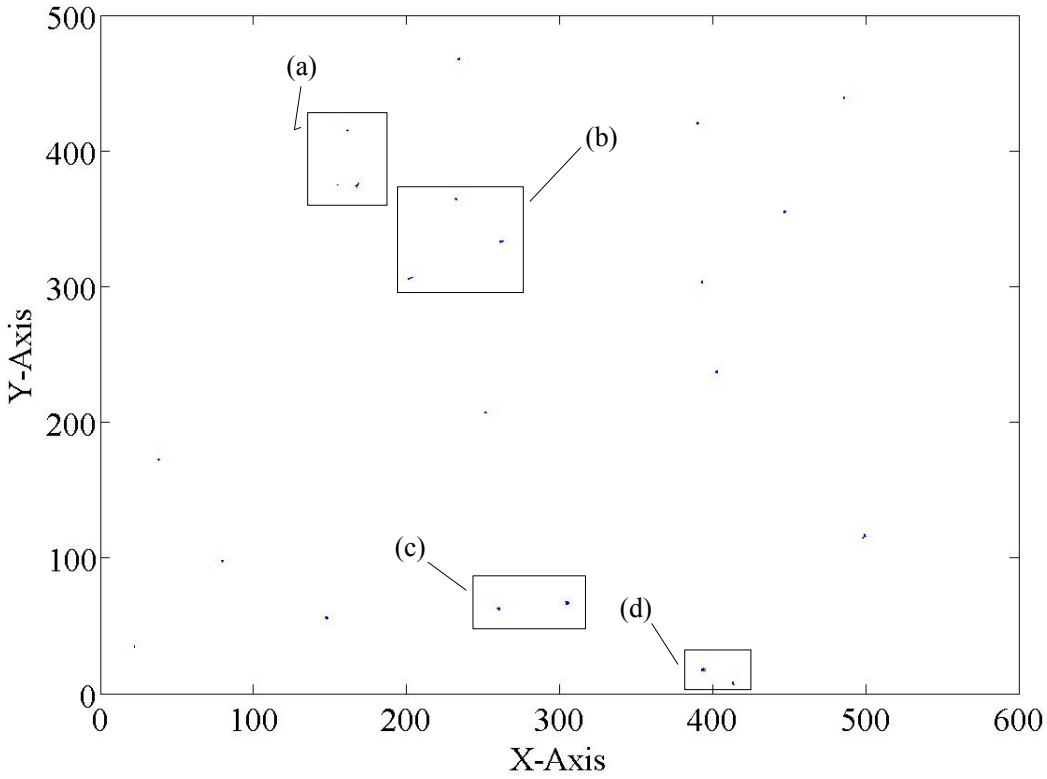


Figure 5.2: Several single transducin trajectories in one plot. X, Y denote the two spatial coordinates. (a), (b), (c) and (d) are denoting areas, that are shown bigger in Figure 5.3. Range: $600 \times 500 \mu m$.

The range of the single trajectories, as plotted in the figures is approx. $1.5 \times 1.5 \mu m$, and the complete area in which the trajectories are distributed has a dimension of $600 \times 500 \mu m$.

The fact that the whole data set is consisting of 5721 data points solely can become problematic during the parameter estimation process, as it is obviously difficult to obtain statistically relevant results for these few data points. This is confirmed by our estimation tests on artificial example trajectories. There we observed that the estimation procedure was much more stable, when we had sufficient data points, some 10000 and up to 100000 data points (see section 3.12 on page 46).

Nevertheless we will try to set up a model and will accomplish the estimation procedure on this data in the next section and discuss the results.

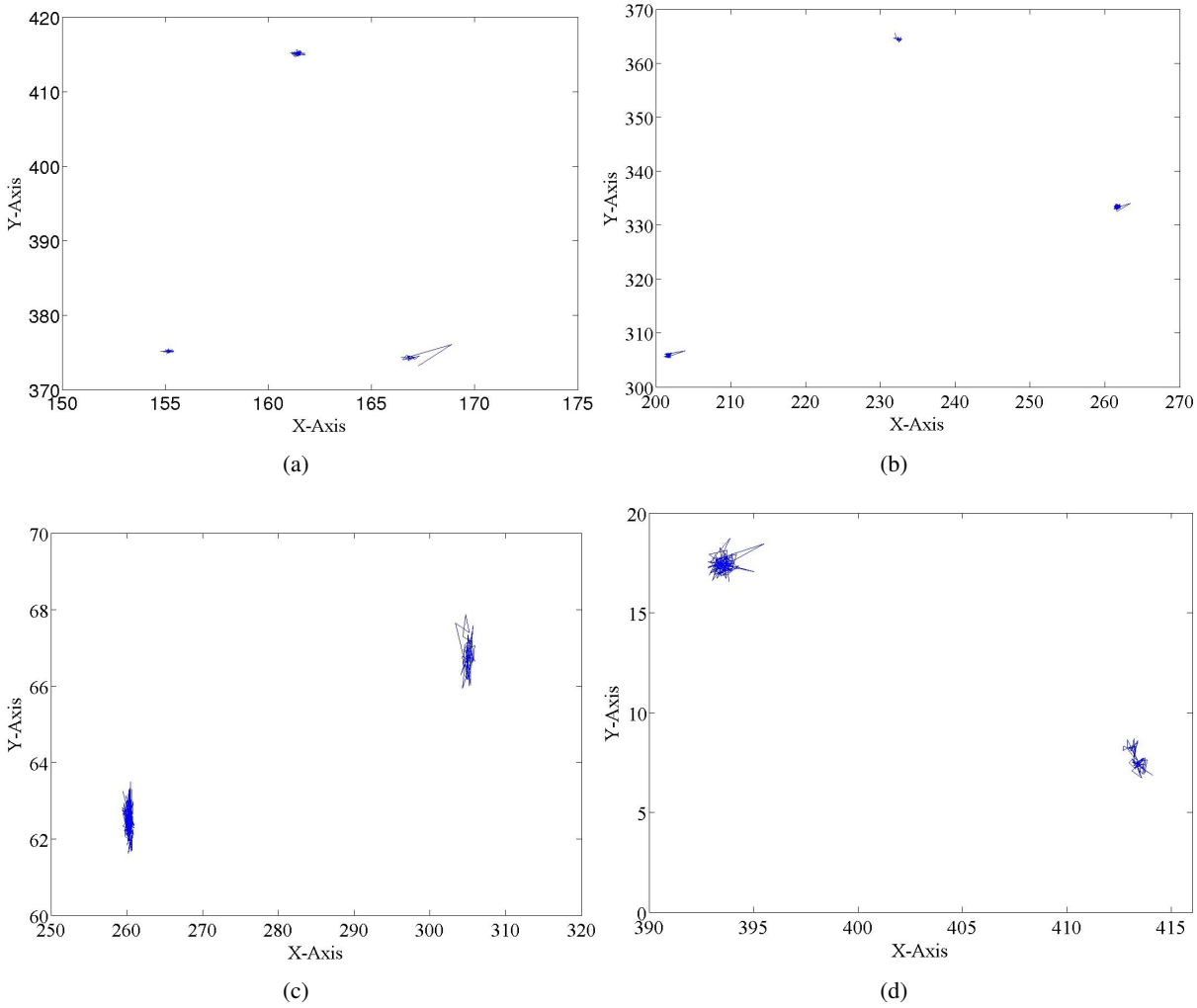


Figure 5.3: Enlargement of the small areas, marked in Figure 5.2. X, Y denote the two spatial coordinates.

5.2 Model Ansatz and Estimation of Parameters

Now that we have figured out the type of experimental data which we want to model, it is convenient to present a model ansatz, upon which the following parameter estimation then should base.

This model ansatz will be based on the Hidden Markov Model - Stochastic Differential Equation (HMM-SDE) concept which was investigated in section 3.10. We want to look for a model which exhibits hidden states of Markovian type, generating output that is describable by stochastic differential equations of Langevin type. The model is characterized by a diffusion model with underlying sequence of hidden Markov states. HMM-SDE is in general represented by a Hidden Markov Model - Vector Auto Regression (HMM-VAR, section 3.11) with a memory depth of one.

While the experimental transducin data is two-dimensional, the model has to be 2-dimensional also. Just in the framework of equation (3.86) we can set up the following diffusion model with Hidden Markov States:

Model 5.1 (Diffusion Model with Hidden Markov States).

$$\dot{\vec{r}} = \Sigma^{(q)} \dot{\vec{W}} \tag{5.1}$$

where $q \in \{1,2,3\}$ *denotes the hidden states and*

$$\begin{aligned} \textit{state space:} \quad & \vec{r} = \begin{pmatrix} x \\ y \end{pmatrix} \\ \textit{noise intensities:} \quad & \Sigma^{(q)} = \begin{pmatrix} \sigma_x^{(q)} & 0 \\ 0 & \sigma_y^{(q)} \end{pmatrix} \\ \textit{Wiener process:} \quad & \vec{W} = \begin{pmatrix} W_x \\ W_y \end{pmatrix} \end{aligned}$$

We have a path of hidden states with length N and three distinct state types. These states are:

1. *The transducin protein is in the solvent and is freely diffusing. The accordant diffusion coefficient is computeable via equation* (3.49) *on page 35:*

$$D^{(1)} = 2k_B^2 T^2 \left(\Sigma^{(1)}\right)^{-1}.$$

2. *The transducin protein is bound to the membrane and performs free diffusion, but with a diffusion coefficient* $D^{(2)}$ *different from* $D^{(1)}$*, that is* $D^{(2)} > D^{(1)}$.

3. *The transducin molecule binds to a rhodopsin or to a complex of rhodopsins and the whole complex is diffusing together on the membrane, although the associated diffusion coefficient will be very tiny, that is* $D^{(3)} \ll D^{(2)} < D^{(1)}$.

Since we now have formulated our model, and given the observation data from the experiments, the estimation process will be accomplished in the framework of HMM-VAR (see section 3.11 on page 43). We already tested the HMM-VAR tools through artificial two-dimensional HMM-SDE data and will accomplish the procedure of parameter estimation on the experimental data in the same framework. The parameter set is then

$$\lambda = \left(\vec{\pi}, \mathscr{T}, \Sigma^{(q)}\right),$$

with the initial distribution $\vec{\pi}$, the transition matrix $\mathscr{T}$ and the set of noise intensities, one per hidden state, $\Sigma^{(q)} = \left(\sigma^{(1)}, \sigma^{(2)}, \sigma^{(3)}\right)$.

Before we present the estimation results, some pre-considerations: As we already mentioned in subsection 3.12.3 on page 56, it is due to numerical reasons not feasible to operate with numerical

values that are lying in the magnitude of 10^{-35}. Also there we pointed out, that rescaling can be a way to overcome this problem, provided that we know in which magnitude the according motion is taking place. In the case of the transducin motion we know that the range of motion is lying in the magnitude of micrometers, thus we can rescale the observation and the according estimated noise intensity parameters with 10^{-35} since diffusion coefficients in the magnitude of micrometers are leading to noise intensities in the magnitude of 10^{-35} (see section 4.6 on page 71).

The estimation delivers the following results for a HMM-SDE model with three hidden states:

$$\begin{aligned}
\vec{\pi}_{est} &= \begin{pmatrix} 0.0 \\ 1.0 \\ 0.0 \end{pmatrix}, \\
\mathscr{T}_{est} &= \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.04 & 0.93 & 0.03 \\ 0.77 & 0.23 & 1.3 \cdot 10^{-58} \end{pmatrix}, \\
\sigma_{est}^{(1)} &= \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix} \cdot 10^{-35}, \\
\sigma_{est}^{(2)} &= \begin{pmatrix} 0.3 & 0.0 \\ 0.0 & 0.35 \end{pmatrix} \cdot 10^{-35}, \\
\sigma_{est}^{(3)} &= \begin{pmatrix} 162.5 & 0.0 \\ 0.0 & 163.4 \end{pmatrix} \cdot 10^{-35}.
\end{aligned}$$

Here we have a probability of one that the hidden state sequence is starting with state two. Furthermore the estimated transition matrix is assigning a holding probability of 0.93 to state two. State one has a holding probability of 0.77, while the holding probability of state three is nearly zero.

The noise intensities are leading to the following three diffusion coefficients

$$\begin{aligned}
D^{(1)} &= \begin{pmatrix} 1.6 & 0.0 \\ 0.0 & 1.6 \end{pmatrix} \frac{\mu m^2}{s}, \\
D^{(2)} &= \begin{pmatrix} 5.4 & 0.0 \\ 0.0 & 5.4 \end{pmatrix} \frac{\mu m^2}{s}, \\
D^{(3)} &= \begin{pmatrix} 0.01 & 0.0 \\ 0.0 & 0.01 \end{pmatrix} \frac{\mu m^2}{s}.
\end{aligned}$$

In section 4.6 on page 71 we stated that our diffusion coefficients for transducin should lie in a range, justified by recent works on the rhodopsin-transducin system, given by (4.3 on page 72)

$$D_T = 0.8 \frac{\mu m^2}{s}, \ldots, 7.0 \frac{\mu m^2}{s}.$$

Thus our estimation results for the first and the second state diffusion coefficient is lying direct within this range, while the third one is too low.

These results for the three diffusion coefficients do fit in the framework of our model partially. The model is basing on the three types of transducin motion: free diffusion in the solvent, free diffusion on the membrane and free diffusion together with rhodopsin (or a complex of rhodopsin molecules). According to these three states it is expected that the first diffusion coefficient should be the highest of the three. This is obviously not the case here.

Another question is the question for the assignment of the three estimated hidden states to the three phases of the transducin motion. Justification can be given physically, as it is expected that transducin molecules should move slower, while diffusing laterally on the membrane, compared to a motion in the solvent above the membrane, and even slower while bound to a rhodopsin or a rhodopsin complex.

In view of this justification the first and second estimated states should be interchanged, meaning that the estimated first diffusion coefficient $D^{(1)} = 1.6\frac{\mu m^2}{s}$ represent the lateral diffusion on the membrane, while the second diffusion coefficient $D^{(2)} = 5.4\frac{\mu m^2}{s}$ is representing the diffusional motion in the solvent. The third state, with a diffusion coefficient of $D^{(3)} = 0.01\frac{\mu m^2}{s}$ in this scope represent the rhodopsin-bound state. Compared to the chosen range for transducin diffusion coefficients (see equation (4.3) on page 72) the first and second estimated coefficient are really good while the third coefficient $D^{(3)}$ is too low. The third coefficient, though it is much too low compared to the expected coefficients (the above mentioned range), shows the correct tendency. This is due to the expectation that the transducin which is bound to a rhodopsin or a complex of rhodopsins should move considerably slower than a free transducin in the solvent or on the membrane.

In the case of three states the transition matrix of the estimated hidden states reveals that the holding probability for the third state is zero. According to the parameter estimation (HMM-VAR) of the observation data a third state is not detectable. That is why we decided to additionally perform parameter estimation based on a diffusion model with two hidden states. Accomplishing HMM-VAR estimation procedure on the data yields the following estimated parameter set

$$\begin{aligned}
\vec{\pi}_{est} &= \begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix}, \\
\mathscr{T}_{est} &= \begin{pmatrix} 0.91 & 0.09 \\ 0.86 & 0.14 \end{pmatrix}, \\
\sigma_{est}^{(1)} &= \begin{pmatrix} 0.6 & 0.0 \\ 0.0 & 0.6 \end{pmatrix} \cdot 10^{-35}, \\
\sigma_{est}^{(2)} &= \begin{pmatrix} 162.1 & 0.0 \\ 0.0 & 159.1 \end{pmatrix} \cdot 10^{-35},
\end{aligned}$$

and leads to the following two diffusion coefficients:

$$\begin{aligned}
D^{(1)} &= \begin{pmatrix} 2.7 & 0.0 \\ 0.0 & 2.7 \end{pmatrix} \frac{\mu m^2}{s}, \\
D^{(2)} &= \begin{pmatrix} 0.01 & 0.0 \\ 0.0 & 0.01 \end{pmatrix} \frac{\mu m^2}{s}.
\end{aligned}$$

Here the first diffusion coefficient is still in the range of diffusion coefficients that we chose. The second state diffusion coefficient is identical to the one obtained in the three-states HMM-SDE model, a circumstance that gives a kind of confirmation to it.

The first coefficient $D^{(1)}$ can be assigned both to the state where the transducin is diffusing in the solvent and to the state where it is diffusing on the membrane. For the second diffusion coefficient we can write the same what we did above, namely that it can be assigned to the state where the transducin molecule is bound to a rhodopsin molecule or even to a complex of rhodopsin molecules. The fact that the considered diffusion coefficient tends to be so low compared to the expected values is an interesting case. The question if on the rod disc membranes the rhodopsin molecules are organized in complexes or if they do not form complexes can possibly be investigated and answered.

The associated transition matrix has holding probabilities $a_{11} = 0.91$ and $a_{22} = 0.14$, while the transition probability from the second state to the first state is relatively high ($a_{21} = 0.81$). Obviously state two is very transient, thus less metastable than state one. In the picture of our transducin motion this circumstance can be interpreted by assigning state two to the rhodopsin-bound case, while state one is a joining-together of the lateral membrane diffusion with the diffusion in the solvent above the membrane, as it was not possible to distinguish the two states through HMM-VAR on the experimental data.

Maybe it is possible to obtain better results when more data points are present: observation of a single transducin for a longer period of time, or by gathering more distinct transducin trajectories with shorter observation time periods.

We have already mentioned that HMM-VAR estimation on artificial diffusion trajectories was precise for 100 000 and more data points. Furthermore we know that tracking of single molecules in the cell is finished when the fluorescence markers are bleached (see section 4.3 on page 65), or when the labeled transducin molecule leaves the area of measurement (see section 4.5 on page 70).

For that reasons mentioned above, it will not be possible to track these molecules for 100 000 frames or more in a row. Instead it will be more feasible to look for a huge ensemble of single transducin trajectories over short periods time.

5.3 Testing the Model

In order to test the fitness of the model, we used the parameter sets that we estimated from the experimental transducin trajectories and generated some artificial observation trajectories with underlying HMM sequences of the length 100.000 based upon these estimated parameters. The trajectory generation was performed just in the way we did in section 3.12 on page 46, where we applied the HMM-VAR parameter estimation on artificial two-dimensional trajectory data. This is meaning that we used the above estimated parameters for the experimental transducin trajectory data as parameters in order to generate artificial diffusion trajectories with underlying sequences of hidden states (we will denote the above estimated parameter set with the nomenclature λ_{gen} while the re-estimated set will be denoted by λ_{est}). Afterwards HMM-VAR has been applied on these trajectories, in order to re-estimate the parameter set, while the initialization of the Baum-Welch

algorithm was done with a Viterbi path, generated by the following transition matrix:

$$\mathscr{T} = \begin{pmatrix} 0.98 & 0.01 & 0.01 \\ 0.01 & 0.98 & 0.01 \\ 0.01 & 0.01 & 0.98 \end{pmatrix}.$$

First we did the estimation for the three states HMM parameter set

$$\begin{aligned} \vec{\pi}_{gen} &= \begin{pmatrix} 0.0 \\ 1.0 \\ 0.0 \end{pmatrix}, \\ \mathscr{T}_{gen} &= \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.04 & 0.93 & 0.03 \\ 0.77 & 0.23 & 1.3 \cdot 10^{-58} \end{pmatrix}, \\ \sigma^{(1)}_{gen} &= \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix} \cdot 10^{-35}, \\ \sigma^{(2)}_{gen} &= \begin{pmatrix} 0.3 & 0.0 \\ 0.0 & 0.3 \end{pmatrix} \cdot 10^{-35}, \\ \sigma^{(3)}_{gen} &= \begin{pmatrix} 163 & 0.0 \\ 0.0 & 163 \end{pmatrix} \cdot 10^{-35}. \end{aligned} \tag{5.2}$$

The re-estimated parameter set then is

$$\begin{aligned} \vec{\pi}_{est} &= \begin{pmatrix} 0.0 \\ 1.0 \\ 0.0 \end{pmatrix}, \\ \mathscr{T}_{est} &= \begin{pmatrix} 0.93 & 0.04 & 0.03 \\ 0.1 & 0.7 & 0.2 \\ 0.23 & 0.77 & 9.1 \cdot 10^{-46} \end{pmatrix}, \\ \sigma^{(1)}_{est} &= \begin{pmatrix} 0.3 & 0.0 \\ 0.0 & 0.3 \end{pmatrix} \cdot 10^{-35}, \\ \sigma^{(2)}_{est} &= \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix} \cdot 10^{-35}, \\ \sigma^{(3)}_{est} &= \begin{pmatrix} 162.8 & 0.0 \\ 0.0 & 163.4 \end{pmatrix} \cdot 10^{-35}. \end{aligned} \tag{5.3}$$

By comparison of 5.3 and 5.2, we observe the following results:

- The transition matrix entries of the re-estimated parameter set for the states 1 and 3 are interchanged.
- The transition probabilities a_{31} and a_{32} in the third row of the re-estimated transition matrix are interchanged.
- The noise intensities $\sigma^{(q)}$ with $q \in \{1,2,3\}$ are estimated correctly.

Running the procedure for a artificial observation trajectory, based upon the same parameters but with the length 6000 delivered the same results.

Now we took the estimated parameter set for the two sates HMM model

$$\begin{aligned}
\vec{\pi}_{gen} &= \begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix}, \\
\mathscr{T}_{gen} &= \begin{pmatrix} 0.91 & 0.09 \\ 0.86 & 0.14 \end{pmatrix}, \\
\sigma^{(1)}_{gen} &= \begin{pmatrix} 0.6 & 0.0 \\ 0.0 & 0.6 \end{pmatrix} \cdot 10^{-35}, \\
\sigma^{(2)}_{gen} &= \begin{pmatrix} 162.1 & 0.0 \\ 0.0 & 159.1 \end{pmatrix} \cdot 10^{-35},
\end{aligned}$$

and after generating artificial trajectory data, based on this parameter set and running estimation procedure on the this trajectory data, we received the following re-estimated parameter set

$$\begin{aligned}
\vec{\pi}_{est} &= \begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix}, \\
\mathscr{T}_{est} &= \begin{pmatrix} 0.91 & 0.09 \\ 0.86 & 0.14 \end{pmatrix}, \\
\sigma^{(1)}_{est} &= \begin{pmatrix} 0.6 & 0.0 \\ 0.0 & 0.6 \end{pmatrix} \cdot 10^{-35}, \\
\sigma^{(2)}_{est} &= \begin{pmatrix} 162.8 & 0.0 \\ 0.0 & 160.1 \end{pmatrix} \cdot 10^{-35}.
\end{aligned}$$

Obviously the results for the two states HMM parameter set are even better than those for the three states HMM parameter set.

5.4 Separate Estimation on the Experimental Transducin X- and Y-Components

Here we present the results of the estimation procedure on the single x- and y-components of the experimental transducin two-dimensional observation data, each regarded separately.

We assume a two-states HMM-SDE model as basis of the estimation process and thus the results for the diffusion coefficients are:

x-component:

$$\begin{aligned}
D^{(1)} &= 3.1 \frac{\mu m^2}{s}, \\
D^{(2)} &= 0.01 \frac{\mu m^2}{s}.
\end{aligned}$$

y-component:

$$\begin{aligned} D^{(1)} &= 3.1\frac{\mu m^2}{s}, \\ D^{(2)} &= 0.01\frac{\mu m^2}{s}. \end{aligned}$$

As we can see the estimated diffusion coefficients for the first state $D^{(1)}$ are also in good agreement to our chosen range for transducin diffusion coefficients, but we did not receive better results for the second state.

Due to this observation we can conclude that an estimation of the single x- and y-components is not really recommended in order to achieve a better estimation results, compared to the estimation on the joint data set. This is an obvious result by the fact that we have not a sufficient amount of data, and a further division of this data in two distinct sets do not achieve a refinement of the estimation results.

5.5 Estimation of the Noise Intensities for Single Trajectories

Here we present the results of the estimation procedure on single trajectories. In the sections 5.2 to 5.4 the estimation was applied to the whole data.

Single trajectories were stored in separate files in order to ensure separate estimation on the single trajectories. Afterwards we took the estimated covariance matrices for each of these trajectories, calculated the variances respectively the noise intensities σ for each of these trajectories, and plotted the results separately for the x- and the y-components of the single trajectories (see Figure 5.4 (a) and (b)).

The average value (mean) of the noise intensity σ_m for both components is

$$\sigma_m = 0.35 \cdot 10^{-35}\,\frac{J \cdot Kg}{s}.$$

Thus the obtained average diffusion coefficient for both components is

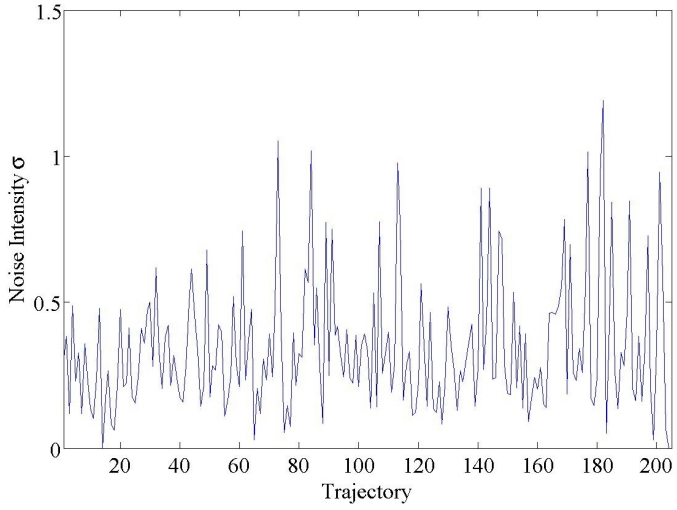
$$D_m = 4.5\,\frac{\mu m^2}{s},$$

which is actually lying also in the chosen transducin diffusion coefficient range

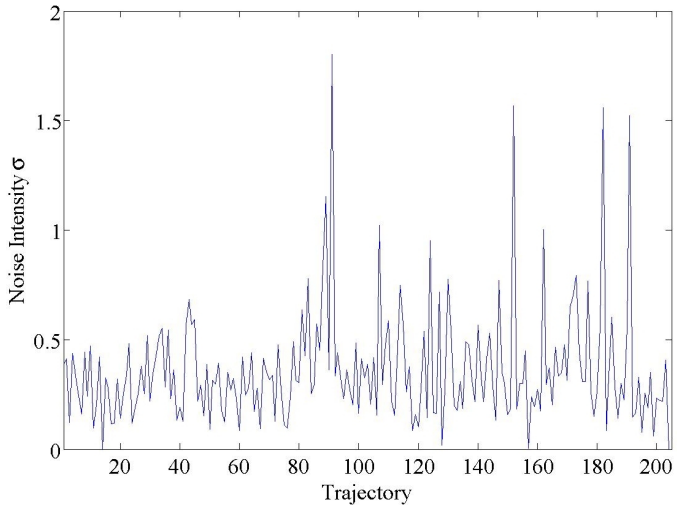
Regarding Figure 5.4 (a) and (b) a certain tendency of different states can be observed, especially on Figure 5.4 (a), which is showing the noise intensity σ_y for the different trajectories. There are some narrow peaks which result in noise intensities between $1.0\,\frac{J \cdot Kg}{s}$ and $1.7\,\frac{J \cdot Kg}{s}$, leading to diffusion coefficients between $1.0\,\frac{\mu m^2}{s}$ to $1.6\,\frac{\mu m^2}{s}$. This circumstance actually reproduces the expectation that transducin diffusion coefficients are lying in a wide range, fluctuating around the mean value D_m. We considered and justified this circumstance in section 4.6 on page 71, with recent works on the topic as background.

Especially Saxton and Owicki ([31]) presented a first work on the concentration dependency of the diffusion coefficients. Due to this our results are successful, although they are because of the

small amount of data statistically not so relevant, as we wish. But this can be improved, when we obtain more observation data.



(a) x-component of the trajectories



(b) y-component of the trajectories

Figure 5.4: Estimation of the noise intensities for single transducin trajectories.

In the next section we will present a theoretical formulation of an HMM-VAR estimation framework that is based on observation data consisting of position differences instead of positions, while the estimation procedure is identical.

5.6 Estimation on the Basis of Differentials d_x in Place of Positions x

Based on the transducin observation data that we have presented in the first section of this chapter (section 5.1), the following considerations can be made. When we consider the single transducin trajectories to exist in two different path of motion as shown in Figure 5.5, namely the red-marked on the one hand and the blue-marked on the other hand, then the two paths are distinguishable through their range of motion.

State one has a small range of motion compared to state two. For the reasons mentioned in the discussion of the last sections it was difficult to distinguish between these two states in the framework of HMM-VAR estimation. One reason is the small amount of data which we have. A second reason is that estimation of the observation data based upon positions in time could perhaps not be an appropriate method.

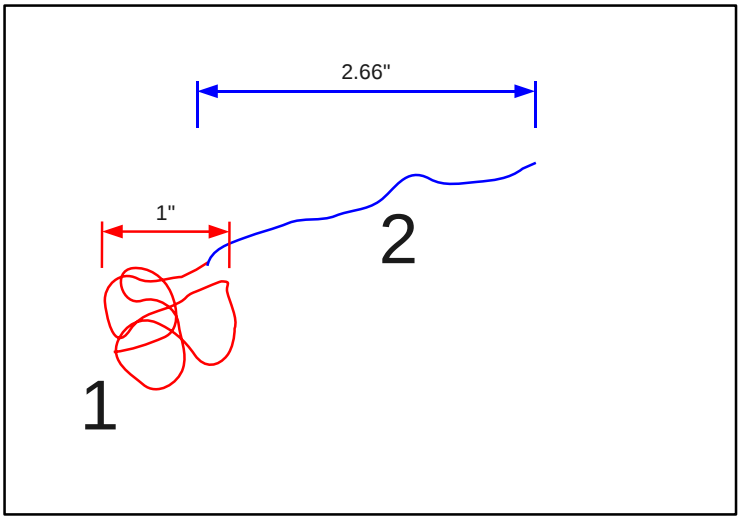


Figure 5.5: Two-States-Assumption. Physically state one is representing a bounded diffusion, while state two is a classical Brownian motion (free diffusion).

However we do not want to leave the framework of HMM-VAR estimation framework, thus an alternative is the formulation of an HMM-SDE model that is based on differentials d_x between two subsequent positions x in time instead of a model which is based on the positions themselves.

In the following a basic theoretical foundation for such a HMM-SDE model is provided, while the implementation in some object oriented programming language like java will not be treated here and is a future task.

Imagine the following situation: We have an observation data that is consisting of positions of an arbitrary particle in time. For the sake of simplicity let it be one-dimensional. Let us now regard two arbitrary positions subsequent in time: $x(t_1)$ and $x(t_2)$. The differential between two positions for a sufficiently small time window can be defined by a limit process

$$d_x = \lim_{(t_2-t_1)\to 0} \left(\frac{x(t_2) - x(t_1)}{t_2 - t_1} \right) = \lim_{\Delta t \to 0} \left(\frac{\Delta x}{\Delta t} \right). \tag{5.4}$$

Furthermore, as we automatically prosecute through this procedure a fragmentation of the trajectory data in discrete differentials, we should also transform from time t as parametrization to the time windows ξ as parametrization. This circumstance is visualized in Figure 5.6.

Thus after this transformation of the problem, from a position state description to a differential state description, we are able to model the new approach through a stochastic differential equation of the following kind

$$\frac{d}{d\xi} d_x = -h(d_x, \xi) + g(d_x, \xi) dW, \tag{5.5}$$

with $g(d_x, \xi)$ and $h(d_x, \xi)$ denoting some functions, depending on d_x and ξ, and with

$$\begin{aligned} h(d_x, \xi) &= 0 \quad , \quad \text{if } d_x \to \infty \text{ for } \Delta t \to 0, \\ h(d_x, \xi) &> 0 \quad , \quad \text{else.} \end{aligned} \tag{5.6}$$

Modeling the function $h(d_x, \xi)$ with a potential gradient, for instance a harmonic potential of the following kind

$$U = \frac{1}{2} F_P \left(d_x - \mu_P \right)^2,$$

with the potential slope F_P and the potential mean μ_P, we get a simple model for observations with two types of motion, like shown in Figure 5.5. In state 1 we have a steady progression of d_x, induced by $h(d_x, \xi) > 0$, while at the boundary from state 1 to state 2, the case should switch to a in-steady progression of d_x, since we have $h(d_x, \xi) = 0$, thus it is $d_x \to \infty$ for $\Delta t \to 0$. Physically state 1 is a bounded diffusion, while state 2 is representing a standard Brownian motion (free diffusion).

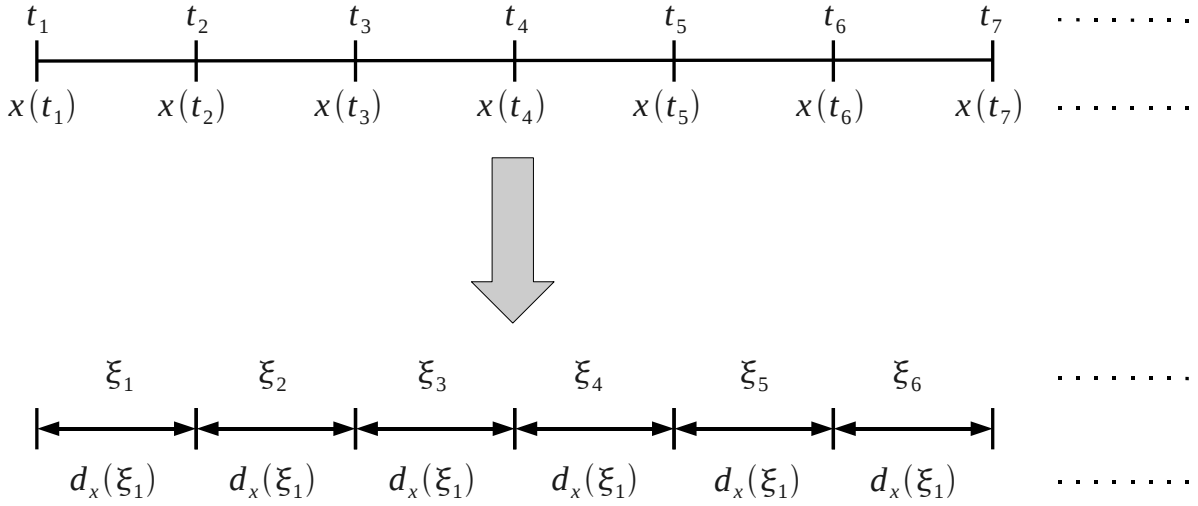


Figure 5.6: Sequence of differentials versus sequence of positions. The Figure has to visualize the idea that will be elaborated here, namely the idea to estimate for parameters of a model on the basis of differentials d_x in time windows ξ in place of positions x in time t.

In section 3.10 on page 37 the theoretical background of parameter estimation procedure based upon a HMM-SDE model was delivered. In order to construct a comparable estimation framework, based upon the model that is presented here, and grounding on equation (5.5) and the two conditions (5.6), we have to construct a Fokker-Planck equation for the SDE (5.5).

We can construct a Fokker-Planck equation that is representing an ensemble picture, an ensemble of trajectories, generated by the SDE (5.5) from the Kramers-Moyal forward expansion. Here we will present the results without general and extensive derivation which is delivered in the appendix (see section A.8 on page 113).

The Kramers-Moyal forward expansion in this case is given by

$$\frac{\partial \rho(d_x,\xi)}{\partial \xi} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial (d_x)} \right)^n \left[D^{(n)}(d_x,\xi) \cdot \rho(d_x,\xi) \right], \tag{5.7}$$

with $D^{(n)}(d_x,\xi)$ giving the Kramers-Moyal coefficients and $\rho(d_x,\xi)$ the probability density function for solutions of the SDE (5.5).

Since the Wiener process that generates the noise term of the SDE is a Gaussian, δ-distributed Langevin force, all Kramers-Moyal coefficients with $n \geq 3$ will be zero and thus we get the Fokker-Planck equation

$$\frac{\partial \rho(d_x,\xi)}{\partial \xi} = \left[-\frac{\partial}{\partial (d_x)} D^{(1)}(d_x,\xi) + \frac{\partial^2}{\partial (d_x)^2} D^{(2)}(d_x,\xi) \right] \cdot \rho(d_x,\xi), \tag{5.8}$$

with the drift- and diffusion coefficients

$$\begin{aligned} D^{(1)}(d_x,\xi) &= h(d_x,\xi)+\frac{\partial g(d_x,\xi)}{\partial(d_x)}g(d_x,\xi), \\ D^{(2)}(d_x,\xi) &= g^2(d_x,\xi). \end{aligned} \tag{5.9}$$

Taking into account a simple SDE like (5.5), with the conditions (5.6) and the following assumptions for the functions $h(d_x,\xi)$ and $g(d_x,\xi)$

$$\begin{aligned} h(d_x,\xi) &= \frac{\partial}{\partial(d_x)}U(d_x,\xi) \\ &= 2F_P(d_x-\mu_P), \\ g(d_x,\xi) &= \sigma, \end{aligned}$$

with σ, denoting the noise intensity, we get for the drift and diffusion coefficients

$$\begin{aligned} D^{(1)}(d_x,\xi) &= 2F(d_x-\mu_P), \\ D^{(2)}(d_x,\xi) &= (\sigma)^2 \\ &=: B. \end{aligned}$$

The following considerations are running just analogous to the Hidden Markov Model case with Stochastic Differential Equation output (HMM-SDE), that was presented in section 3.10 on page 37, with the exception that the observation is not $x(t)$ but $d_x(\xi)$.

Thus we can write down a solution ansatz for the Fokker-Planck equation (5.8)

$$\rho(d_x,\xi) = A(\xi)\exp\left(-(d_x-\mu(\xi))\cdot\Sigma\cdot(d_x-\mu(\xi))^T\right),$$

where $\mu(\xi)$ denotes the mean of the random variable $d_x(\xi)$ and $\Sigma = \left(\sigma^{(1)},\sigma^{(2)}\right)$ the noise intensities for the two states.

This solution ansatz leads to the following three ordinary differential equations (ODEs) for the first state (bounded diffusion)

$$\begin{aligned} \frac{d}{d\xi}\mu &= -F_P\left(d_x-\mu_P\right), \\ \frac{d}{d\xi}\Sigma &= -2F_P\Sigma-2B\Sigma^2, \\ \frac{d}{d\xi}A &= (F_P-B\Sigma)A, \end{aligned}$$

and three ODEs for the second state (free diffusion)

$$\begin{aligned}\frac{d}{d\xi}\mu &= 0\\ \frac{d}{d\xi}\Sigma &= -2B\Sigma^2,\\ \frac{d}{d\xi}A &= -AB\Sigma.\end{aligned}$$

Assuming a HMM on the basis of these two states leads to consideration of two additional parameters, namely the vector of the initial Markov chain distribution $\vec{\pi}$ and the two states transition matrix $\mathscr{T}$. The other parameters are the potential mean μ_P, the potential slope F_P and the noise intensities B. Thus we have to estimate the following two sets of parameters:

$$\begin{aligned}\lambda^{(1)} &= \{\vec{\pi}, \mathscr{T}, \mu_P, F_P, B\},\\ \lambda^{(2)} &= \{\vec{\pi}, \mathscr{T}, B\}.\end{aligned}$$

The two ODE systems above can be solved analogously to the HMM-SDE case for $x(t)$ as observation and the reinsertion of these solutions in the solution statement $\rho(d_x, \xi)$ enables us to formulate the complete data likelihood of the observation O and the sequence of a two-states HMM q, given an arbitrary parameter set λ

$$\mathscr{L}(\lambda) = \mathscr{L}(\lambda|O,q) = p(O,q|\lambda),$$

with

$$p(O,q|\lambda) = \pi(q_0)\rho(O_0|q_0)\prod_{\xi=1}^{\Gamma}\mathscr{T}(q_{\xi-1},q_\xi)\rho(O_\xi|q_\xi,O_{\xi-1}).$$

This is obviously analogous to the HMM-SDE section 3.10, while the observation here is consisting of a sequence of d_x in ξ.

The further development of the estimation on the basis of these contemplations runs also analogous to the HMM-SDE estimation procedure with $x(t)$ as observation by definition of forward- and backward variables α_ξ, β_ξ and subsequent derivation of the maximum estimators $\hat{\lambda}^{(q)}$ ($q \in \{1,2\}$).

Afterwards an analogous formulation in the framework of HMM-VAR (Hidden Markov Model - Vector Auto Regression) can be tackled, when estimation of observation data with memory depth higher than one is desired.

6 Conclusion and Outlook

The main aspect of this thesis was to derive a physical model for the diffusion processes involved in the first steps of vision transduction from single molecule experimental data. To this end, we combined the framework of Hidden Markov Models with simple stochastic differential equations (HMM-SDE).

After introduction of all desired physical and mathematical concepts a model describing the molecular diffusion was delivered. With the help of the Hidden Markov Model - Vector Auto Regression (HMM-VAR) method with a memory depth one which is in fact a HMM with a memorless overdamped Langevin equation (HMM-SDE), we estimated the transition rates and the diffusivities for this model.

The model was successfully applied to model data. Here we generated artificial trajectory data, both two- and three-dimensional, based on the HMM-SDE model with predefined parameter sets. The application of estimation procedure on the two-dimensional artificial data was successful, provided that the length of the considered HMM sequences were sufficiently large (some 10 000 to more than 100 000 observation points). The estimation procedure was able to detect the underlying parameter sets.

In the scope of this thesis we have decided to choose for the transducin molecule a diffusion coefficient in a certain range (see also section 4.6 on page 71), given by

$$D_T = (0.8, \ldots, 7.0)\,\frac{\mu m^2}{s}.$$

This range is based on several recent investigations on the rhodopsin-transducin system (in the appendix we have presented a selected number of works on the topic, see section A.2 on page 101).

The pure estimation process on the available experimental data in the framework of HMM-VAR was not entirely successful due to the small amount of data available. A second difficulty was the tendency of HMM-VAR to get stuck in local maxima of the Likelihood landscape. However, this latter problem should be at least partially compensated when more experimental data is available.

Nevertheless the model estimation from experimental data was partially successful. We estimated a three-states HMM and received at least two diffusion coefficients which are in the above mentioned range. The diffusion coefficient we calculated for the first state of our three-states HMM model is $D^{(1)} = 1.6\,\frac{\mu m^2}{s}$ and for the second state $D^{(2)} = 5.4\,\frac{\mu m^2}{s}$. This is of course a nice result, although the remaining diffusion coefficient for the third state is a bit too far away from the known range.

The same holds for the two-states HMM model. The estimated diffusion coefficient for the first state was with $D^{(1)} = 2.7\,\frac{\mu m^2}{s}$ in good agreement with known diffusion coefficient values. Again, the estimated coefficient for the second state was far out of the expected range.

The mean noise intensity over all trajectories leads to a mean diffusion coefficient of $D_m = 4.5\,\frac{\mu m^2}{s}$ which also is lying in the above range. Furthermore the plots of the considered noise intensities versus the single trajectories are showing a wide variance around this mean value. Some of the trajectories produce noise intensities that are leading to diffusion coefficients in the range of $1.0\,\frac{\mu m^2}{s}$ to $1.6\,\frac{\mu m^2}{s}$. This suggests that the transducin interactions on the disc membrane may be more complex than expected and may require more than three states whose biophysical interpretation is yet unclear. This calls for further research in collaboration with the experimentalists.

In order to guide future studies, we have proposed a theoretical formulation of an HMM-VAR estimation framework that is based on observation data consisting of position differences instead of positions, while the estimation procedure is identical. We anticipate that this new method will compensate for some of the deficiencies observed with the methods used in this work.

Some options for future studies are performed in the following:

1. We have to obtain more experimental trajectory data. This will allow us to accomplish better estimation of diffusion constants and also a better determination of the associated hidden states in the framework of HMM-VAR. Our confidence to reach this aim is justified by the estimation results that we have conducted on artificial trajectory data and presented in section 3.12 on page 46. The good performance of HMM-VAR is shown there, provided that the considered observation data has a sufficient large length (some 10 000 and more than 100 000 observation points).
2. The use of global optimization methods like the genetic and the simulated annealing algorithm (see section 3.13 on page 57) is the next step. The application of this algorithms on existing and additional data may allow better estimation results.
3. Finally a direct simulation of the considered biological process in the framework of a general cell simulation has to be aproached. In our research group an implementation of a so called ”Cellular Molecular Dynamics” (CMD) has been provided. The CMD is a first approach in order to simulate an arbitrary cell with a solvent and arbitrary definable particles which are diffusing in the solvent or on the membrane. Furthermore one is able to define a type of reaction between these particles. The simulation of the retinal rod disc membranes with integral proteins rhodopsin and the transducin proteins that are diffusing in the solvent between and on the rod disc membranes should be possible in the framework of the CMD.

7 Bibliography

[1] R. C. Aster, B. Borchers, and C. H. Thurber: *Parameter Estimation and Inverse Problems* (Elsevier Academic Press, 2005), 1st ed.

[2] N. J. T. Bailey: *The Elements of Stochastic Processes - with Applications to the Natural Sciences* (John Wiley & Sons, 1964), 1st ed.

[3] G. P. Basharin, A. N. Langville, and V. A. Naumov: *The Life and Work of A. A. Markov*, Numerical Solution of Markov Chains (Workshop) pp. 1–22 (1990).

[4] H. C. Berg: *Random Walks in Biology* (Princeton University Press, 1993), new, expanded ed.

[5] J. A. Bilmes: *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models* (1998).

[6] F. Bruckert, M. Chabre, and T. M. Vuong: *Kinetic Analysis of the Activation of Transducin by photoexcited Rhodopsin (Influence of the lateral Diffusion of Transducin and Competition of GDP and GTP Nucleotide Site)*, Biophysical Journal **63**, 616 (1992).

[7] M. E. BURNS and T. D. LAMB: *Visual Transduction by Rod and Cone Photoreceptors*, in *Chalupa LM, Werner LS, editors. The Visual Neurosciences. USA. Bradford Books, MIT Press; 2003. pp. 215-233. chapter 16.*.

[8] S. Doose and M. Sauer: *Fluoreszenz bringt Licht ins Dunkel - Fluoreszenzmikroskopie ermöglicht das Studium einzelner Moleküle auch in lebenden Zellen*, Physik Journal 6 (2007) Nr. 12, S. 55-60 (2007).

[9] A. Einstein: *Investigations on the Theory of the Brownian Movement* (Dover Publications, Inc., 1956), 2nd ed.

[10] A. Einstein: *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Annalen der Physik pp. 549–560 (1905).

[11] S. Felber, H. Breuer, F. Petruccione, J. Honerkamp, and K. Hofmann: *Stochastic Simulation of the Transducin GTPase Cycle*, Biophysical Journal **71**, 3051 (1996).

[12] W. Feller: *An Introduction to Probabilty Theory and its Applications - Volume I* (John Wiley & Sons, 1968), 3rd ed.

[13] W. Feller: *An Introduction to Probabilty Theory and its Applications - Volume II* (John Wiley & Sons, 1968), 3rd ed.

[14] J. Herivel: *The Man and the Physicist* (Oxford University Press, 1975).

[15] S. Heyse, O. P. Ernst, Z. Dienes, K. P. Hofmann, and H. Vogel: *Incorporation of Rhodopsin in Laterally Structured Supported Membranes: Observation of Transducin Activation with Spatially and Time-Resolved Surface Plasmon Resonance*, Biochemistry **37**, 507 (1998).

[16] I. Horenko, E. Dittmer, A. Fischer, and C. Schütte: *Automated Model Reduction for Complex Systems exhibiting Metastability*, Mult. Mod. Sim., 5 (3). pp. 802-827. Official URL: http://dx.doi.org/10.1137/050623310 (2005).

[17] I. Horenko, E. Dittmer, and C. Schütte: *Reduced Stochastic Models for Complex Molecular Systems*, Comp. Vis. Sci., 9 (2). pp. 89-102. Official URL: http://dx.doi.org/10.1007/s00791-006-0021-1 (2005).

[18] C. S. I. Horenko, C. Hartmann and F. Noé: *Data-Based Parameter Estimation of Generalized Multidimensional Langevin Processes*, Phys. Rev. E, 76:016706 (2007).

[19] M. Kahlert and K. P. Hofmann: *Reaction Rate and Collisional Efficiency of the Rhodopsin-Transducin System in Intact Retinal Rods*, Biophysical Journal **59**, 375 (1991).

[20] E. Fermi: *Thermodynamics* (Dover Publications, Inc. New York, 1937 and 1956), 2nd ed.

[21] M. Fischbach: *Methods for Modeling Metastable Conformational Dynamics from Trajectory Data, Diploma Thesis, Department of Mathematics and Computer Science Freie Universität Berlin, year = 2008,*.

[22] D. C. Lamb and C. Bäuchle: *Dem Molekül auf der Spur - Mithilfe eines Weitfeld-Fluoreszenzmikroskops lässt sich der Weg einzelner Moleküle und Teilchen verfolgen*, Physik Journal 6 (2007) Nr. 12, S. 39-45 (2007).

[23] T. D. Lamb: *Gain and Kinetics of Activation in the G-Protein Cascade of Phototransduction*, Proc. Natl. Acad. Sci. USA **93**, 566 (1996).

[24] E. Meerbach and C. Schütte: *Sequential Change Point Detection in Molecular Dynamics Trajectories*, Preprint submitted to Elsevier, to be published (2008).

[25] R. Nelson: *Probability, Stochastic Processes, and Queuing Theory - The Mathematics of Computer Performance Modelling* (Springer, 1995), 1st ed.

[26] B. Oksendal: *Stochastic Differential Equations* (Springer, 2007), 6th ed.

[27] L. Pachter and B. Sturmfels: *Algebraic Statistics for Computational Biology* (Cambridge University Press, 2005).

[28] I. Prigogine and I. Stengers: *Dialog mit der Natur - Neue Wege naturwissenschaftlichen Denkens* (Piper, Neuausgabe 1990), 6th ed.

[29] L. R. Rabiner: *A Tutorial on Hidden Markov Models and selected Applications in Speech Recognition* pp. 267–296 (1990).

[30] H. Risken: *The Fokker-Planck Equation* (Springer, 1989), 2nd ed.

[31] M. J. Saxton and J. C. Owicki: *Concentration Effects on Reactions in Membranes: Rhodopsin and Transducin*, Biochimica et Biophysica Acta **979**, 27 (1989).

[32] Scheibe and Erhard: *Die Philosophie der Physiker*, vol. 1 (C.H. Beck Verlag, 2006).

[33] C. Schütte and I. Horenko: *Likelihood-Based Estimation of Multidimensional Langevin Models and its Application to Biomolecular Dynamics, Multiscale Modelling and Simulation*, to appear (2007).

[34] F. Schwabl: *Statistische Mechanik* (Springer-Verlag, 2006), 3rd ed.

[35] P. D. P. Schwille: *Einblicke ins Leben - Mit den Methoden der Nanobiophotonik die kleinsten Werkstätten des Lebens entschlüsseln*, Physik Journal 6 (2007) Nr. 12, S. 35-38 (2007).

[36] C. U. M. Smith: *Elements of Molecular Neurobiology* (John Wiley & Sons, LTD, 2002), 3rd ed.

[37] M. von Smoluchowski: *Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen*, Bullet. Int. Crac. pp. 756–780 (1906).

[38] A. Tarantola: *Inverse Problems Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics, Philadelphia, 2005), 1st ed.

[39] P. Schlüsche: *Untersuchung der Funktion und Dynamik von DNA-Transkriptionsfaktoren mittels Einzelmolekül-F]luoreszenzmikroskopie, PHD Thesis, school = Fakultat für Chemie und Pharmazie der Ludwig-Maximilians-Universität München, year = 2007*.

[40] T. Funatsu, Y. Harada, M. Takanaga, K. Saito, and T. Yanagida: *Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution*, Nature 374 (1995), Nr. 6522 pp. 555–559 (1995).

[41] A. Ishijima and T. Yanagida: *Single molecule nanobioscience*, Trends Biochem Sci 26 (2001), Nr. 7 pp. 438–444 (2001).

[42] D. S. Talaga: *Markov processes in single molecule fluorescence*, Colloid and Interface Science 12 (2007) pp. 285–296 (2007).

[43] D. Axelrod: *Total Internal Reflection Fluorescence Microscopy*, in *D. Taylor and Y. Wang: Methods in Cell Biology, Vol. 30, New York* (Academic Press, 1989), pp. 245–270.

[44] R. Brown: *A Brief Account on Microscopical Observations on the Particles Contained in the Pollen of Plants and on the General Existence of Active Molecules in Organic and Inorganic Bodies*, in *Brown, Robert. 1866. The miscellaneous botanical works of Robert Brown: Volume 1. (Edited by John J. Bennett). R. Hardwicke, London. and later: Mabberley, D.J. 1985. Jupiter Botanicus: Robert Brown of the British Museum. Lubrecht and Cramer Ltd, London.* (1828).

[45] C. J. Needham, J. R. Bradford, A. J. Bulpitt, and D. R. Westhead: *A Primer on Learning in Bayesian Networks for Computational Biology*, PLoS Computational Biology August 2007 Volume 3 Issue 8 e129 pp. 1409–1416 (2007), URL `www.ploscompbiol.org`.

[46] T. Weise: *Global Optimization Algorithms, Theory and Application* (2009), URL `http://www.it-weise.de`.

[47] P. M. Pardalos and e. Mauricio G. C. Resende: *Handbook of Applied Optimization.* (Oxford University Press, Inc., 198 Madison Avenue, New York, New York 10016, USA,).

[48] I. Rechenberg.: *Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* (Friedrich Frommann Verlag, Stuttgart, 1973.).

[49] J. H. Holland: *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence* (MIT Press, Cambridge, MA, USA, 1992).

[50] D. E. Goldberg: *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989).

[51] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, , and E. Teller.: *Equation of state calculations by fast computing machines.*, The journal of Chemical Physics **21**, 1087 (1953).

[52] S. Kirkpatrick, C. D. G. Jr., and M. P. Vecchi.: *Optimization by Simulated Annealing.*, Science **220**, 671 (1983).

[53] A. Nolte and R. Schrader: *A note on the finite time behaviour of simulated annealing*, Mathematics of Operations Research **347**, 476 (2000).

[54] P. J. M. van Laarhoven and e. E. H. L. Aarts: *Simulated Annealing: Theory and Applications* (Springer, Kluwer Academic Publishers, Norwell, MA, USA, June 30, 1987, 1987).

[55] B. Hajek: *Cooling schedules for optimal annealing*, Mathematics of Operations Research, (13) (1988).

[56] M. Schöngen: *Rasterkraftmikroskopie und Fluoreszenzspektroskopie zur nichtinvasiven und hochaufgelösten Untersuchung von Nanostrukturen, Diplomarbeit, Institut für Optik und Atomare Physik, Technische Universität Berlin* (2008).

[57] J. L. E. Dreyer: *Tycho Brahe. A picture of scientific life and work in the sixteenth century, Edinburgh, 1890*.

[58] V. Westphal and S. W. Hell: *Nanoscale Resolution in the Focal Plane of an Optical Microscope*, Phys. Rev. Lett. **94**, 143903 (2005).

[59] S. W. Hell: *Far-Field Optical Nanoscopy*, Science **316**, 1153 (2007), URL `http://www.sciencemag.org/cgi/content/abstract/316/5828/1153`.

[60] H. Nyquist: *Thermal Agitation of Electric Charge in Conductors*, Phys. Rev. **32**, 110 (1928).

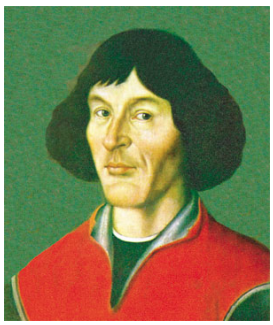
[61] H. B. Callen and T. A. Welton: *Irreversibility and Generalized Noise*, Phys. Rev. **83**, 34 (1951).

[62] R. Kubo: *Statistical-Mechanical Theory of Irreversible Processes. I: General Theory and Simple Applications to Magnetic and Conduction Problems*, Journal of the Physical Society of Japan **12**, 570 (1957), URL `http://jpsj.ipap.jp/link?JPSJ/12/570/`.

Index

A Appendix

A.1 From Copernicus to Newton - A Historical Example for the Role of Observation on Modeling



Nicolaus Copernicus

It can be seen as the birth date of the modern science era, when Nicolaus Copernicus (* 19th February 1473 in Thorn; † 24th May 1543 in Frauenburg) based on observations of himself and others, without modern instruments like telescopes, developed a model of the planetary motion around the sun. His work that took him centuries of his life, and published by him at the year of his death, constituted the heliocentric system in contradiction to the geocentric system. In the geocentric system the earth was thought as the center of the world orbited by the sun and the planets, while in the heliocentric system the sun is thought as the center that is orbited by the planets. Though the heliocentric system already was introduced by the Greek natural philosopher Aristarchus of Samos, it was Copernicus who initially provided a mathematical model it.

Since the basis of the model by Copernicus is the circular path of the planets around the sun, his model possesses quantitative failures. Later Kepler delivered a model with his famous three rules grounding on elliptic orbits, and for this reason is in a better agreement with observation.

In the comparison between the two models on the same problem becomes manifest the role of observation on model development. It was an extraordinary lucky coincidence for Johannes Kepler to have the observation data that was accumulated by Tycho Brahe.

Tycho Brahe

Tycho Brahe, born Tyge Ottesen Brahe (* 14th December 1546 in Schonen; † 24th October 1601 in Prague) was a danish nobleman, and maybe one of the most outstanding astronomer of all times. One has to keep in mind that Brahe was the last great astronomer of the pre-telescope era, the era in which positions of planets and fixed stars were measured by instruments such as the quadrant. Nevertheless Brahe was able to accumulate data of such precision (accurate to the arc minute, or 1/30 the width of the full moon) and such quantity that his achievement for the modern sciences is immense.

More or less one can say that Brahe was the prototype of a scientific observer. The accuracy of his observations were tremendous for his lifetime and he has influenced the ideal of the modern science that was obeyed by the following generations of scientists, namely

the ideal of continuous accuracy in measurement and observation.

The huge amount of observation data that Brahe has accumulated were published later by Kepler in the so called ”Rudolphine Tables”, named after Emperor Rudolf II. This position catalog of planets and fixed stars is forming the basis on which both Kepler’s model of planetary motion and Isaac Newton’s cosmological theory ground.

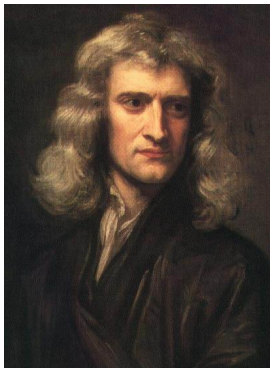
Johannes Kepler

Johannes Kepler (also: Ioannes Keplerus; * 27th December 1571 in Weil der Stadt; † 15th November 1630 in Regensburg) inherited this huge amount of data, and in contrary to Brahe who only had the aim to accumulate more and more precise data, Kepler had the aim to give these data a sense. Here at last comes the role of the theoretician. Kepler who was a mathematician, astronomer and natural philosopher was such a theoretician, and in contrary to Brahe also believed in the truth of the heliocentric system.

Kepler had the data from Brahe but the difficulty was to give sense to them, and it took him nearly twenty years of trial and examination to find his three rules of planetary motion around the sun.

The difficulty was the fact that the orbiting paths of the planets are observed from the earth, while the earth itself is orbiting around the sun. This can lead to the apparently queer observation that the planets are performing loops. By taking the position data of mars and the sun as basis, and without knowing the correct orbit form of mars, Kepler inversely derived the positions of the earth and from this the orbit of mars.

It was the notion of the force action at a distance (”Anima Motrix”) together with the examination of Brahe’s data which brought Kepler to the discovery that the mars orbit is not a circle but an ellipse. This is not obvious since the elliptic paths of the big planets are nearly circular. That he anyway came to this insight is also due to Tycho Brahe and the extraordinary accuracy of his data which allowed a distinction between a circular and an elliptic path.



Isaac Newton (Portrait by Godfrey Kneller from 1689)

Finally Kepler concluded that the planetary orbits are of elliptic form while the sun is in one of the two foci (1st Kepler law of planetary motion). The second proposition of his model reveals that a line joining a planet and the sun sweeps out equal areas during equal intervals of time (2nd Kepler law of planetary motion). Kepler published these two laws in the year 1609 in his work ”Astronomia Nova”.

Later Newton formulated in his work ”Philosophiae Naturalis Principia Mathematica” (1687) the universal theory of classical mechanics that combines the mechanical phenomena of normal day life with the laws of gravitation from which also at least the laws of Kepler are following.

A.2 Important Papers on the Rhodopsin-Transducin Process

There are numerous papers, which are dealing with the analysis of the rhodopsin-transducin visual transduction pathway (R-T-system):

1. One of the early works on this topic is served by Michael J. Saxton and John C. Owicki in a paper from 1989 ([31]). Saxton and Owicki are investigating the question, how the concentration of the numerous proteins affect the reaction rate of the rhodopsin-transducin-system. They are setting up a steady-state-diffusion model with concentration dependent diffusion coefficients. They do this with the assumption of lateral diffusing transducin proteins on the membrane, and finally point out that there exists a maximum (optimal) reaction rate, and along the way an optimal concentration. This conclusion grounds on the fact that on the one hand the reaction rate increases by the law of mass action, and on the other hand a higher concentration would decrease the rate due to the decrease of the diffusion coefficients.

2. A work from 1991 is [19], in which the authors deliver a model of transducin activation from its partial reactions, and solve the kinetic equations of the model both numerically, and for small photo-activation, analytically.

3. One year later, in the year 1992, a paper was published [6], in which the authors investigate the activation of transducin by photo-excited rhodopsin in terms of the Michaelis-Menten-enzymology and additionally by experiments.

4. Lamb published his findings in a paper from 1996 [23], and in this work refers about a simulation of the diffusional interaction of the proteins by Monte Carlo techniques. His conclusion is that there exists a linear activation of many transducins by only one activated rhodopsin.

5. [15] is an experimental work from 1998, giving an account on the observation of transducin activation with spatially and time resolved surface plasmon resonance.

6. [11] deliver Master equation simulations of the diffusional protein interaction on a rectangular $1\mu m^2$ model membrane divided into 15×15 cells and with mono- and biomolecular reactions within cells and diffusional transitions between (neighboring) cells.

A.3 Probability Theory

Definition A.1 (Probability Measure).

A probability measure P on the sample space Ω is a function of subsets of Ω satisfying three axioms:

1. *For every set $A \subset \Omega$, the value of the function is a nonnegative number:*
   $P(A) \geq 0$.
2. *For any two disjoint sets A and B, the value of the function for their union $A+B$ is equal to the sum of its value for A and B seperately:*
   $P(A+B) = P(A) + P(B)$, *provided that* $AB = \varnothing$.
3. *Axiom of countable additivity:*
   For a countably infinite collection of disjoint sets A_k, $k = 1,2,\ldots$, we have
   $P(\sum_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$.
4. *The value of the function for Ω (as a subset) is equal to 1:*
   $P(\Omega) = 1$.

Definition A.2 (Random Variable).

A numerically valued function X of ω with domain Ω:
$\omega \in \Omega : \ \omega \rightarrow X(\omega)$
is called a ***random variable***.

Definition A.3 (Mathematical Expectation).

For a random variable X defined on a countable sample space Ω, its mathematical expectation is the number $E(X)$, given by the formula:
$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$,
provided that the series converges absolutely, namely
$\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) < \infty$.
In this case we say that the mathematical expectation of X exists.

Definition A.4 (Continuous valued Random Variables and Mathematical Expectation with densities).

Consider a function f defined on $\mathbb{R}^1 = (-\infty, +\infty)$:
$u \to f(u)$ with $u \in \mathbb{R}^1$,
and satisfying the two conditions:

1. $u \to f(u) \geq 0$,
2. $\int_{-\infty}^{+\infty} f(u)du = 1$.

Such a function is called a density function on $\mathbb{R}^1$.
We are now able to define also the probabilities for any continuous valued random variable X by means of a density function f so that for any closed interval $[a,b]$ we have
$P(a \leq X \leq b) = \int_a^b f(u)du$.
More generally, if A is the union of intervals not necessarily disjoint and some of which may be infinite, we have
$P(X \in A) = \int_A f(u)du$.
The mathematical expectation for X and any arbitrary function of it, $\Phi(X)$, then is defined as follows:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} f(u)du, \\ E(\Phi(X)) &= \int_{-\infty}^{+\infty} \Phi(u)f(u)du. \end{aligned}$$

Definition A.5 (Conditional Probability).

Let $A, B \subset \Omega$ be two events. Then $P(A|B)$ is denoted as the probability of event A such that B has occured and is defined as:

$P(A|B) = \frac{P(A \cap B)}{P(B)}$, *if* $P(B) \neq 0$.

Definition A.6 (Independent Random Variables).

The countably valued random variables $X_1,\ldots,X_n$ are said to be independent if for any sequence of real numbers $x_1,\ldots,x_n$ we have

$P(X_1 = x_1,\ldots,X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2)\cdots P(X_n = x_n)$.

A.4 Definitions for Optimization

The following definitions are all taken from [46].

Definition A.7 (Objective Function).

An objective function $f : \mathbb{X} \mapsto \mathbb{Y}$ with $\mathbb{Y} \subseteq \mathbb{R}$ is a mathematical function, which is subject to optimization.

Definition A.8 (Problem Space).

The problem space $\mathbb{X}$ (phenome) of an optimization problem is the set containing all elements x which could be the solution of the optimization problem.

Definition A.9 (Local Maximum).

A (local) maximum $\hat{x}_l \in \mathbb{X}$ of one (objective) function $f : \mathbb{X} \mapsto \mathbb{R}$ is an input element with $f(\hat{x}_l) \geq f(x)$ for all x neighboring $\hat{x}_l$.

Definition A.10 (Local Minimum).

A (local) minimum $\check{x}_l \in \mathbb{X}$ of one (objective) function $f : \mathbb{X} \mapsto \mathbb{R}$ is an input element with $f(\check{x}_l) \leq f(x)$ for all x neighboring $\check{x}_l$.

Definition A.11 (Local Optimum).

A (local) optimum $x_l^* \in \mathbb{X}$ of one (objective) function $f : \mathbb{X} \mapsto \mathbb{R}$ is either a local maximum or a local minimum.

Definition A.12 (Global Maximum).

A global maximum $\hat{x} \in \mathbb{X}$ of one (objective) function $f : \mathbb{X} \mapsto \mathbb{R}$ is an input element with $f(\hat{x}) \geq f(x)$ for all $x \in \mathbb{X}$.

Definition A.13 (Global Minimum).

A global minimum $\check{x} \in \mathbb{X}$ of one (objective) function $f : \mathbb{X} \mapsto \mathbb{R}$ is an input element with $f(\check{x}) \leq f(x)$ for all $x \in \mathbb{X}$.

Definition A.14 (Global Optimum).

A global optimum $x^* \in \mathbb{X}$ of one (objective) function $f : \mathbb{X} \mapsto \mathbb{R}$ is either a global maximum or a global minimum.

Definition A.15 (Optimal Set).

The optimal set X^* is the set that contains all optimal elements.

Definition A.16 (Solution Candidate).

A solution candidate x is an element of the problem space $\mathbb{X}$ of a certain optimization problem.

Definition A.17 (Solution Space).

We call the union of all solutions of an optimization problem its solution space $\mathbb{S}$.

$$X^* \subseteq \mathbb{S} \subseteq \mathbb{X} \tag{A.1}$$

Definition A.18 (Search Space).

The search space $\mathbb{G}$ of an optimization problem is the set of all elements g which can be processed by the search operations.

Definition A.19 (Genotype).

The elements $g \in \mathbb{G}$ of the search space $\mathbb{G}$ of a given optimization problem are called the genotypes.

Definition A.20 (Gene).

The distinguishable units of information in a genotype that encode the phenotypical properties are called genes.

Definition A.21 (Search Operations).

The search operations searchOp are used by optimization algorithms in order to explore the search space $\mathbb{G}$.

Definition A.22 (Genotype-Phenotype Mapping).

The genotype-phenotype mapping (GPM) $gpm : \mathbb{G} \mapsto \mathbb{X}$ is a left-total binary relation which maps the elements of the search space $\mathbb{G}$ to elements in the problem space $\mathbb{X}$.

$$\forall g \in \mathbb{G} \quad \exists x \in \mathbb{X} : gpm(g) = x \tag{A.2}$$

Definition A.23 (Optimization Problem).

An optimization problem is defined by a fivetuple ($\mathbb{X}$, F, $\mathbb{G}$, Op, gpm) specifying the problem space $\mathbb{X}$, the objective functions F, the search space $\mathbb{G}$, the set of search operations Op, and the genotype-phenotype mapping gpm. In theory, such an optimization problem can always be solved if Op is complete and the gpm is surjective.

Definition A.24 (Completeness).

A set Op of search operations $searchOp$ is complete if and only if every point g_1 in the search space $\mathbb{G}$ can be reached from every other point $g_2 \in \mathbb{G}$ by applying only operations $searchOp \in Op$.

Definition A.25 (Individual).

An individual p is a tuple $(p.g, p.x)$ of an element $p.g$ in the search space $\mathbb{G}$ and the corresponding element $p.x = gpm(p.g)$ in the problem space $\mathbb{X}$.

Definition A.26 (Population).

A population Pop is a list of individuals used during an optimization process.

$$Pop \subseteq \mathbb{G} \times \mathbb{X} : \forall p = (p.g, p.x) \in Pop \Rightarrow p.x = gpm(p.g) \tag{A.3}$$

A.5 The Auto-Correlation Function and its Spectral Representation[1]

Let $x(t)$ be an arbitrary random variable, then the auto-correlation function, which is a measure for the dependence of the random variable x at different times, is given by

$$C_{x,x}(t_1,t_2) = \langle x(t_1)x(t_2)\rangle = \int x_1 x_2 P(x_1,t_1;x_2,t_2)dx_1 dx_2 \tag{A.4}$$

with $t_2 = t_1 + \tau > t_1$ and $P(x_1,t_1;x_2,t_2)dx_1dx_2$. It denotes the joint probability that $x(t)$ is inside the closed interval $[x_1, x_1 + dx_1]$ at time t_1, and conditionally within the closed interval $[x_2, x_2 + dx_2]$ at time t_2.

This correlation function can be expressed through the transition probability $P(x_2,t_2|x_1,t_1)$, that is

$$P(x_1,t_1;x_2,t_2) = P(x_2,t_2|x_1,t_1)P(x_1,t_1). \tag{A.5}$$

In equilibrium, the fluctuations of the variable $x(t)$ form a stationary process, meaning that $P(x,t)$ is independent of the time t and the transition probability density depends on the time difference $\tau = t_2 - t_1$ solely.

Thus the auto-correlation function is independent of the running time, that is

$$C_{x,x}(\tau) = \int \left(x_1 P(x_1) \int x_2 P(x_2,\tau|x_1)dx_2 \right) dx_1. \tag{A.6}$$

The second integral is the conditional mean value of the random variable $x(t)$, thus

$$C_{x,x}(\tau) = \int x_1 \langle x(\tau)\rangle_{x(0)=x} P(x_1)dx_1. \tag{A.7}$$

For a multivariate stationary process $x_i(t)$ with $i = 1,\ldots,n$, the correlation functions form a matrix with elements

$$C_{i,j}(\tau) = \int x_i \langle x_j(\tau)\rangle_{x_j(0)=x_j} P(x_1,\ldots,x_n)dx_1\cdots dx_n \tag{A.8}$$

with the associated Fourier-transform

$$S_{i,j}(\omega) = \int_{-\infty}^{+\infty} C_{i,j}(\tau)e^{i\omega\tau}d\tau, \tag{A.9}$$

called the "power spectrum" of the fluctuating values for the random variable $x(t)$.

A.6 The Langevin Equation as a First Order System

The classical Langevin equation that we presented in this thesis (see subsection 3.9.1 on page 32) can also be presented in form of a first order system for the canonical coorinates q_k and momenta p_k, with $k \in \{1,\ldots,N\}$ and $N \in \mathbb{N}$.

[1] "Lectures on the Stochastic Dynamics of Reacting Biomolecules", lecture notes by W. Ebeling, Yu. Romanovsky and L. Schimansky-Geier (eds.); Humboldt University 2002; download-url: http://www.physik.hu-berlin.de/studium/vlv/materialien/skripten_html

This is given by

Definition A.27 (Langevin equation as a first Order System).

$$\dot{q}_k = \tfrac{1}{m} p_k, \tag{A.10}$$

$$\dot{p}_k = -\nabla U(\{q_k\}) - \tfrac{\gamma}{m} P_k + \sigma \dot{W}(t), \tag{A.11}$$

with W denoting the Wiener process.

With $p_k = m v_k$, derivation of (A.10) one more time with respect to time t and insertion of (A.11) in it leads to

$$\begin{aligned}
\ddot{q}_k(t) &= \frac{1}{m}\dot{p}_k(t), \\
\iff \ddot{q}_k(t) &= \frac{1}{m}\left[-\nabla U(\{q_k\}) - \frac{\gamma}{m} p_k(t) + \sigma \dot{W}(t)\right], \\
\iff \ddot{q}_k(t) &= -\frac{1}{m}\nabla U(\{q_k\}) - \frac{\gamma}{m^2} p_k(t) + \frac{\sigma}{m}\dot{W}(t), \\
\iff \dot{v}_k(t) &= -\frac{1}{m}\nabla U(\{q_k\}) - \frac{\gamma}{m^2} m v_k(t) + \frac{\sigma}{m}\dot{W}(t), \\
\iff m\dot{v}_k(t) &= -\nabla U(\{q_k\}) - \frac{\gamma}{m} v_k(t) + \sigma \dot{W}(t).
\end{aligned}$$

This is finally the classical general Langevin equation we already know.

A.7 The Fluctuation-Dissipation-Theorem

The Fluctuation-Dissipation-Theorem (FDT) is maybe one of the most universal relations in mathematical physics, relating both in classical statistical physics and quantum theory the reversible fluctuations with the irreversible dissipation of energy into heat within a system.

Historically, Einstein and Smoluchowski first derived a FDT from their considerations on the phenomenon of Brownian motion. The Einstein-Smoluchowski-Relation correlates the diffusion coefficient D (depending on the fluctuations) of a Brownian particle, suspended in water, with its mobility μ (depending on the dissipation)

$$D = \mu k_B T. \tag{A.12}$$

A general FDT was formulated first in the year 1928 by Harry Nyquist as an explanation for the "Johnson–Nyquist noise" [60] which is the electronic noise, generated by the thermal agitation of the charge carriers (usually the electrons) inside an electrical conductor at equilibrium, occurring

regardless of any applied voltage. Here the FDT is given by the mean squared voltage $\langle V^2 \rangle$ in relation to the resistance R:

$$\langle V^2 \rangle = 4Rk_BT\,\Delta\nu, \tag{A.13}$$

while $\Delta\nu$ denote the bandwidth over which the voltage is measured.

But a prove to general FDTs was initially delivered in 1951 by Herbert B. Callen and Theodore A. Welton [61].

In the year 1957 the Japanese mathematical physicist Ryogo Kubo provided an ingenious piece of work [62] in which he presented a groundbreaking derivation of FDT in the framework of the so called "Linear Response Theory" (LRT), also developed by him. He is stating in the abstract:

> "A general type of fluctuation-dissipation-theorem is discussed to show that the physical quantities such as complex susceptibility of magnetic or electric polarization and complex conductivity for electric conduction are rigorously expressed in terms of time-fluctuation of dynamical variables associated with such irreversible processes. This is a generalization of statistical mechanics which affords exact formulation as the basis of calculation of such irreversible quantities from atomistic theory. The general formalism of the statistical-mechanical theory is examined in detail. The response, relaxation, and correlation functions are defined in quantum mechanical way and their relations are investigated. The formalism is illustrated by simple examples of magnetic and conduction problems. Certain sum rules are discussed for these examples. Finally it is pointed out that this theory may be looked as a generalization of the Einstein relation".

Kubo initially introduces the notion of the linear response function, before he derives the general FDT from it.

He applies a simple but elegant motivation of the liner response of a system due to a perturbation for Hamiltonian systems. Kubo's considerations are so ingenious that we should present them simply here, just in the way he is introducing it in his seminal paper from 1957.

Consider an isolated system, described by a Hamiltonian H_0 which is the energy of the system, externally unperturbed. Suppose furthermore an external force field $f(t)$ that is switched on at a definite time and disturbing the system. This external force field will lead to a new Hamiltonian

$$H = H_0 - xf(t).$$

The new quantity $x(t)$ that is introduced here, denote an arbitrary observable of the system that we consider here.

In the following, Kubo makes the assumption of weak perturbation that allows it to take into account linear approximation of the whole case. The aim is to derive a relation for the response of the system on this weak perturbation. One has to consider now an arbitrary physical quantity, denoted by $y(t)$. Then we obtain the response of the system through the change of this quantity, given by Δy.

Kubo makes preliminary a treatment in the scope of classical mechanics by introducing a statistical ensemble with the associated initial distribution function $\rho_0(p_k, q_k)$ in the phase space of the

related system. Then the dynamical motion of the unperturbed system is given by the Liouville equation:

$$\frac{\partial \rho_0}{\partial t} = -\sum_k \left(\frac{\partial \rho_0}{\partial q_k}\frac{\partial H_0}{\partial p_k} - \frac{\partial \rho_0}{\partial p_k}\frac{\partial H_0}{\partial q_k} \right) = \{H_0, \rho_0\}. \tag{A.14}$$

We get this equation by building the total differential of the distribution function $\rho_0(p_k, q_k)$ and assuming this total differential to be zero, what is nothing else than the classical Liouville assumption that for Hamiltonian systems the phase space volume is conserved.

Here the q_k, p_k are giving the set of the canonical momenta and coordinates, and $\{\cdot,\cdot\}$ is denoting the Poisson bracket, defined for two arbitrary phase space functions $A(q_k, p_k)$ and $B(q_k, p_k)$ in the following manner:

$$\{A,B\} = \sum_k \left(\frac{\partial A}{\partial q_k}\frac{\partial B}{\partial p_k} - \frac{\partial A}{\partial p_k}\frac{\partial B}{\partial q_k} \right).$$

When the perturbation $H_p = -xf(t)$ is inserted adiabatically at $t = -\infty$, the perturbed distribution function obeys the following general time evolution equation:

$$\frac{\partial \rho_p}{\partial t} = \{H_0, \rho_p\} + \{H_p(t), \rho_0\}. \tag{A.15}$$

The assumption of linear approximation leads to

$$\rho_p = \rho_0 + \Delta\rho,$$

and further with the assumption of equilibrium which is meaning

$$\frac{\partial \rho_0}{\partial t} = \{H_0, \rho_0\} \stackrel{!}{=} 0,$$

to

$$\frac{\partial (\Delta\rho)}{\partial t} = \{H_0, \Delta\rho\} - f(t)\{x, \rho_0\}. \tag{A.16}$$

Here the time evolution of the perturbance $\Delta\rho$ is connected to unperturbed quantities like the Hamiltonian H_0 of the unperturbed system and the unperturbed density function ρ_0. The solution of this evolution equation is easily to get and is given by

$$\Delta\rho = -\int_{-\infty}^{t} dt' e^{i(t-t')\mathscr{L}} \{x, \rho_0\} f(t'),$$

with $-\infty < t' < t$ and the Liouville operator $\mathscr{L}$, defined through

$$i\mathscr{L}\bullet = \{H, \bullet\}.$$

Thus the change Δy of the dynamical quantity $y(t)$ is given by

$$\begin{aligned} \Delta y(t) &= \int d\Gamma \Delta\rho\, y(p,q) \\ &= -\int d\Gamma \int_{-\infty}^{t} dt' \left[e^{i(t-t')\mathscr{L}} \{x, \rho_0\} \right] f(t') y(p,q) \\ &= -\int d\Gamma \int_{-\infty}^{t} dt' \{x, \rho_0\}\, y(t-t')\, f(t'), \end{aligned} \tag{A.17}$$

with $d\Gamma$, the volume-element of the phase space.

Furthermore it is

$$e^{i(t-t')\mathscr{L}} \cdot y(p,q) = y(t-t').$$

The equation (A.17) means that the response Δy is a superposition of the effects of pulses $f(t')dt'$, with $-\infty < t' < t$. The response for a unit pulse, Kubo denotes by the response function or the after-effect function $\Phi_{yx}(t)$. From equation (A.17) he then concludes

$$\Phi_{yx}(t) = -\int d\Gamma \, \{x, \rho_0\} \, y(t), \tag{A.18}$$

leading for the response Δy to

$$\Delta y(t) = \int_{-\infty}^{t} dt' \, \Phi_{yx}(t-t') \, f(t'). \tag{A.19}$$

The functions $\Phi_{yx}(t-t')$ are the so called response functions with the Fourier transforms of them in the frequency space (with $\tau = t - t'$) given by

$$\Psi_{yx}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_{yx}(\tau) e^{i\omega\tau} d\tau. \tag{A.20}$$

Applying the considerations up to now on an arbitrary random variable $x(t)$ will lead to the general FDT for classical stochastic processes. Without thorough derivation we sketch the case here solely:

The auto-correlation for the random variable $x(t)$ is given by

$$C_{x,x}(\tau) = k_B T \int_{\tau}^{\infty} \Phi(s) ds. \tag{A.21}$$

Derivation with respect to time leads to

$$\frac{\partial C_{x,x}(\tau)}{\partial \tau} = -k_B T \Phi(\tau) \tag{A.22}$$

which is the first form of the FDT. The left hand side is representing the fluctuation, while the right hand side is giving the dissipation.

Introduction of the Fourier-transforms finally delivers the second form of the FDT, that is

$$S_{x,x}(\omega) = \frac{2k_B T}{\omega} \mathscr{I}\left[\Psi(\omega)\right]. \tag{A.23}$$

Here $S_{x,x}(\omega)$, as already presented in section A.5, gives the power spectrum of the correlation function, representing the fluctuation part, and $\mathscr{I}\left[\Psi(\omega)\right]$ denote the imaginary part of the response function Fourier-transform, representing the dissipation in the system.

The FDT can be generalized in a straight-forward way to the case of space-dependent fields, to the case of several variables or to a quantum-mechanics setting.

A.8 Kramers-Moyal Forward Expansion

In the following we will derive the Kramers-Moyal forward expansion and neglect the associated backward equation. Eventually one can show that the solutions of both are equivalent. [30, p. 63]

From the definition of the transition probability follows that the probability density $\rho(x,t+\tau)$ at time $t+\tau$ and the probability density $\rho(x,t)$ at time t are connected by ($\tau \geq 0$)

$$\rho(x,t+\tau) = \int P(x,t+\tau|x',t)\rho(x,t)dx', \tag{A.24}$$

while $P(x,t+\tau|x',t)$ denote the transition probabilities from states x' at times t to states x at times $t+\tau$.

Lets now assume that we know all moments ($n \geq 1$):

$$\begin{aligned} M_n(x',t,\tau) &= \langle [\Gamma(t+\tau)-\Gamma(t)]^n \rangle|_{\Gamma(t)=x'} \\ &= \int (x-x')^n P(x,t+\tau|x',t)dx, \end{aligned} \tag{A.25}$$

where $|_{\Gamma(t)=x'}$ means that at time t the random variable Γ has the sharp value x'.

There are three ways to derive a general expansion of the transition probability. We follow one way, while the two other ways can be seen in [30, p. 63].

If all the moments are given, the characteristic function can be constructed as follows:

$$\begin{aligned} C(u,x',t,\tau) &= \int_{-\infty}^{\infty} e^{iu(x-x')} P(x,t+\tau|x',t)dx \\ &= \sum_{n=0}^{\infty} \frac{(iu)^n}{n!} \int (x-x')^n P(x,t+\tau|x',t)dx \\ &\overset{(A.25)}{=} \sum_{n=0}^{\infty} \frac{(iu)^n}{n!} M_n(x',t,\tau) \\ &= 1+\sum_{n=1}^{\infty} \frac{(iu)^n}{n!} M_n(x',t,\tau) \end{aligned} \tag{A.26}$$

Because the characteristic function is the Fourier transform of the probability density and vice verse we can express the transition probability by the moments M_n

$$\begin{aligned} P(x,t+\tau|x',t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iu(x-x')} C(u,x',t,\tau)du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iu(x-x')} \left[1+\sum_{n=1}^{\infty} \frac{(iu)^n}{n!} M_n(x',t,\tau) \right] \end{aligned} \tag{A.27}$$

With ($n \geq 0$)

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}(iu)^n e^{-iu(x-x')}du = \left(-\frac{\partial}{\partial x}\right)^n \delta(x-x') \tag{A.28}$$

and

$$\delta(x-x')f(x') = \delta(x-x')f(x) \tag{A.29}$$

we have

$$P(x,t+\tau|x',t) = \left[1+\sum_{n=1}^{\infty}\frac{1}{n!}\left(-\frac{\partial}{\partial x}\right)^n M_n(x',t,\tau)\right]\delta(x-x') \tag{A.30}$$

Inserting (A.30) in (A.24) leads to

$$\begin{aligned}\rho(x,t+\tau)-\rho(x,t) &= \frac{\partial\rho(x,t)}{\partial t}\tau + o(\tau^2)\\ &= \sum_{n=1}^{\infty}\frac{1}{n!}\left(-\frac{\partial}{\partial x}\right)^n \int M_n(x',t,\tau)\delta(x-x')\rho(x',t)dx'\\ &= \sum_{n=1}^{\infty}\left(-\frac{\partial}{\partial x}\right)^n\left[\frac{M_n(x,t,\tau)}{n!}\right]\rho(x,t)\end{aligned} \tag{A.31}$$

Let us further assume that the moments M_n can be expanded into a Taylor series with respect to τ ($n \geq 1$):

$$\frac{M_n(x,t,\tau)}{n!} = D^{(n)}(x,t)\tau + o\left(\tau^2\right) \tag{A.32}$$

The term with τ^0 must vanish, because for $\tau = 0$ the transition probability P has the initial value

$$P(x,t|x',t) = \delta(x-x'), \tag{A.33}$$

which leads to vanishing moments in (A.25).

By taking into account only the linear terms in τ we finally get the *"Kramers-Moyal forward expansion"* after using equation (A.32) in (A.31):

$$\boxed{\frac{\partial\rho(x,t)}{\partial t} = \mathscr{L}_{KM} * \rho(x,t),} \tag{A.34}$$

with

$$\boxed{\mathscr{L}_{KM} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^n D^{(n)}(x,t).} \tag{A.35}$$

which is the Kramers-Moyal forward operator.

The transition probability $P(x,t|x',t)$ is the distribution $\rho(x,t)$ for the special initial condition $\rho(x',t) = \delta(x-x')$. Thus the transition probability must also obey equation (A.34), i.e.,

$$\frac{\partial P(x,t|x',t')}{\partial t} = \mathscr{L}_{KM} * P(x,t|x',t'), \tag{A.36}$$

where the initial condition of P is given by (A.33) with t replaced by t'.

A.9 Deriving the Fokker-Planck Equation from the Kramers-Moyal Forward Expansion

In order to show that the Kramers-Moyal expansion (A.34) leads to the Fokker-Planck equation when we take into account Langevin equations with Gaussian, δ-distributed Langevin forces, we first introduce the general Langevin equation for one stochastic variable X(t):

$$\dot{X}(t) = h(X,t) + g(X,t)\Gamma(t), \tag{A.37}$$

The Langevin force $\Gamma(t)$ is again assumed to be a Gaussian random variable with zero mean and δ-distributed correlation function:

$$\langle \Gamma(t) \rangle = 0, \tag{A.38}$$

$$\langle \Gamma(t)\Gamma(t') \rangle = 2\delta(t-t'). \tag{A.39}$$

In general [30, p. 63] the Kramers-Moyal coefficients are given by

$$D^{(n)}(x,t) = \frac{1}{n!} \lim_{\tau \to 0} \frac{1}{\tau} \langle [X(t+\tau) - x]^n \rangle |_{X(t)=x}. \tag{A.40}$$

where $X(t+\tau)$ ($\tau > 0$) is a solution of equation (A.37) which at time t has the sharp value $X(t) = x$.

In order to derive the Kramers-Moyal coefficients we first have to write down (A.37) in integral form:

$$X(t+\tau) - x = \int_t^{t+\tau} [h(X(t'),t') + g(X(t'),t')\Gamma(t')]dt' \tag{A.41}$$

Next we expand the functions h and g with respect to X:

$$h(X(t'),t') = h(x,t') + h'(x,t')[X(t') - x] + \cdots$$
$$g(X(t'),t') = h(x,t') + g'(x,t')[X(t') - x] + \cdots$$

Inserting these in equation (A.41) leads to

$$\begin{aligned} X(t+\tau) - x &= \int_t^{t+\tau} h(x,t')dt' + \int_t^{t+\tau} h'(x,t')[X(t') - x]dt' \\ &\quad + \cdots \\ &\quad + \int_t^{t+\tau} g(x,t')\Gamma(t')dt' + \int_t^{t+\tau} g'(x,t')\Gamma(t')[X(t') - x]dt' \\ &\quad + \cdots \end{aligned} \tag{A.42}$$

Iterating equation (A.42) for $[X(t') - x]$ leads to

$$\begin{aligned} X(t+\tau) - x &= \int_t^{t+\tau} h(x,t')dt' + \int_t^{t+\tau} h'(x,t') \int_t^{t'} h(x,t'')dt''dt' \\ &\quad + \int_t^{t+\tau} h'(x,t') \int_t^{t'} g(x,t'')\Gamma(t'')dt''dt' + \cdots \\ &\quad \int_t^{t+\tau} g(x,t')\Gamma(t')dt' + \int_t^{t+\tau} g'(x,t')\Gamma(t') \int_t^{t'} h(x,t'')dt''dt' \\ &\quad + \int_t^{t+\tau} g'(x,t')\Gamma(t') \int_t^{t'} g(x,t'')\Gamma(t'')dt''dt' + \cdots \end{aligned} \tag{A.43}$$

By repeated iterations only Langevin forces and the known functions g(x,t) and h(x,t) and their derivatives appear on the right-hand side of (A.43).

If we now take the average of (A.43) and use (A.38) and (A.39) we come to

$$\begin{aligned} \langle X(t+\tau) - x \rangle = \textstyle\int_t^{t+\tau} h(x,t')dt' + &\int_t^{t+\tau} \int_t^{t'} h'(x,t')h(x,t'')dt''dt' + \cdots \\ &+ \int_t^{t+\tau} \int_t^{t'} g'(x,t')g(x,t'')2\delta(t'' - t')dt''dt' \\ &+ \cdots \end{aligned} \tag{A.44}$$

For the δ-function in (A.44) we can use any representation $\delta_\varepsilon(t)$ symmetric around the origin:

$$\delta_\varepsilon(t) = \begin{cases} \frac{1}{\varepsilon}, & -\frac{\varepsilon}{2} < t < \frac{\varepsilon}{2} \\ 0, & elsewhere \end{cases} \tag{A.45}$$

and finally when we take $\varepsilon \to 0$ we conclude

$$\begin{aligned} \int_t^{t'} g(x,t'')2\delta(t''-t')dt'' &= g(x,t)\int_t^{t'} 2\delta(t''-t')dt'' \\ &= g(x,t) \end{aligned} \tag{A.46}$$

and with this for (A.44)

$$\begin{aligned} \langle X(t+\tau)-x\rangle = \int_t^{t+\tau} h(x,t')dt' + &\int_t^{t+\tau}\int_t^{t'} h'(x,t')h(x,t'')dt''dt' + \cdots \\ &+ \int_t^{t+\tau}\int_t^{t'} g'(x,t')g(x,t')dt' + \cdots \end{aligned} \tag{A.47}$$

In the limit $\tau \to 0$ we get the first Kramers-Moyal coefficient

$$D^{(1)}(x,t) = h(x,t) + g'(x,t)g(x,t) \tag{A.48}$$

Here we should state that only the lowest terms entered in equation (A.43), because higher terms will vanish for the limit in (A.40). Integrals of the form

$$\langle \int_t^{t+\tau} \cdots \Gamma(t_1) \int_t^{t_1} \cdots \Gamma(t_2) \int_t^{t_2} \cdots \Gamma(t_3) \int_t^{t_3} \cdots \Gamma(t_4) dt_1 dt_2 dt_3 dt_4 \rangle$$

can only lead to results proportional to τ^2 which vanishes for the limit in (A.40). Integrals not containing the Langevin-force are proportional to τ^n, where n is the number of integrals. For the limit in (A.40) thus only one such integral will survive.

Using these arguments we obtain for the second coefficient

$$\begin{aligned} D^{(2)}(x,t) &= \frac{1}{2}\lim_{\tau\to 0}\int_t^{t+\tau}\int_t^{t'} g(x,t')g(x,t'')2\delta(t'-t'')dt'dt'' \\ &= g^2(x,t). \end{aligned} \tag{A.49}$$

By using these arguments for the higher coefficients $D^{(n)}$ we conclude that they all vanish for $n \geq 3$. Thus the final result is

$$\begin{aligned} D^{(1)}(x,t) &= h(x,t) + \frac{\partial g(x,t)}{\partial x} g(x,t) \\ D^{(2)}(x,t) &= g^2(x,t) \\ D^{(n)}(x,t) &= 0 \ \ for\ n \geq 3 \end{aligned} \tag{A.50}$$

In addition to the deterministic drift h(x,t), $D^{(1)}$ contains a term which is called the spurious drift or the noise-induced drift

$$D^{(1)}_{noise-ind} = \frac{\partial g(x,t)}{\partial x} g(x,t) = \frac{1}{2}\frac{\partial}{\partial x} D^{(2)}(x,t) \tag{A.51}$$

It stems from the fact that during a change of $\Gamma(t)$ also X(t) changes and therefore $\langle g(X(t),t)\Gamma(t)\rangle$ is no longer zero.

So finally it is clear that for SDEs of Langevin-type like equation (A.37) with Gaussian noise the Kramers-Moyal forward expansion turns to the Fokker-Planck equation

$$\boxed{\frac{\partial \rho(x,t)}{\partial t} = \mathscr{L}_{FP} * \rho(x,t)} \tag{A.52}$$

where

$$\boxed{\mathscr{L}_{FP} = -\frac{\partial}{\partial x} D^{(1)}(x,t) + \frac{\partial^2}{\partial x^2} D^{(2)}(x,t)} \tag{A.53}$$

is the Fokker-Planck operator.

Danksagungen

Ich möchte folgenden Menschen danken, die mir beistanden, als ich diese Arbeit geschrieben habe aber auch schon vorher auf verschiedenen Wegen es ermöglicht haben, dass ich überhaupt soweit gekommen bin, meine Diplomarbeit anzufangen:

- Prof. Eckehard Schöll dafür, dass er mir die Möglichkeit gegeben hat, diese Arbeit über ein wunderbares Thema zu schreiben.
- Dr. Frank Noé, der ein toller Betreuer war und mich tatkräftig unterstützt hat. Dabei ist er in der Lage, Ernst und Freude an der Wissenschaft zu vermitteln.
- Prof. Christoph Schütte, der rigorose Mathematik mit einem Talent, Verständnis und Freude an dieser Mathematik zu vermitteln, in sich vereinigt.
- Prof. Erwin Sedlmayr, der mir in der Zeit, wo ich auf Grund eines lebensbedrohlichen Unfalls körperliche Probleme hatte, bürokratische Hindernisse aus dem Weg geräumt hat, und mir dadurch überhaupt erst ermöglicht hat, im Studium so schnell wieder auf die Beine zu kommen.
- Prof. Andreas Knorr, der der beste Lehrer für theoretischen Physik ist, den man sich vorstellen kann. Ohne seine Vorlesungen hätte ich nicht diesen Spaß an der theoretischen Physik, das ich jetzt besitze.
- Mein Zimmergenosse Martin Fischbach, den man als einen akribischen Wissenschaftler bezeichnen darf, immer auf der Suche nach Perfektion. Eine Art, welche mich angetrieben hat, eine immer bessere Arbeit zu schreiben. Darüber hinaus ist er jemand, der für jeden Spaß zu haben ist und allein dadurch motiviert er seine Freunde.
- Mein sehr guter Freund Thorsten Lüdge, mit dem man philosophisch-wissenschaftlich immer reden kann und dabei auch schöne Ideen entstehen. Darüber hinaus ist er einer der besten Freunde, die man haben kann
- Franz Josef Schmitt, Ekkehard Ehrenstein und Max Schöngen die tolle Freunde, Wissenschaftler und einfach hilfsbereite Menschen sind.
- Alle Leute, die das Atomic Café bereichern und viele von Ihnen zu besten Freunden geworden sind.
- Martin Held, Jan Wigger, Kawe Yoccef, Sabine Pilari, Axel Raack, Tim Conrad, Illia Horenko, Burkhardt Schmitt, Carsten Hartmann, ... und allen in der BioComp Gruppe, die ich vergessen habe, für eine schöne Zeit.

Darüber hinaus habe ich meinen Eltern sehr viel zu verdanken, ohne deren tagtäglichen Kampf für meine und meiner Geschwister Bildung, ich nicht soweit gekommen wäre. Was ich geschafft habe, ist auch ihr Verdienst.

Auch möchte ich danken meiner Familie insgesammt und vielen vielen Freunden, welche ich hier nicht alle aufführen kann, die aber an meiner Seite waren, als ich auf Grund des oben erwähnten Unfalls, sprichwörtlich mein Leben wieder von Vorne anfangen musste. Sie alle haben mich wieder zurück ins Leben geholfen, indem sie mir gezeigt haben, dass ich sehr sehr vielen Menschen fehlen würde, wenn ich nicht mehr dagewesen wäre.

Nicht zuletzt möchte ich meiner wunderbaren Freundin, Elli Schlottmann danken:

Du hast mir sehr oft dadurch geholfen, dass du einfach für mich da warst, wenn es mir schlecht ging, aber auch wenn es mir gut ging. Dafür liebe ich dich.

Affirmation

Hereby I, Arash Azhand, declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at Technische Universität Berlin or Freie Universität Berlin or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Ich, Arash Azhand, erkläre hiermit an Eides statt, dass ich die vorliegende Diplomarbeit sebständig und ohne unerlaubte Hilfe angefertigt, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Date .....

Signature .....