Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 94

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation
Institute of Complex Systems
Jülich Centre for Neutron Science
Peter Grünberg Institute

Lecture Notes of the
46th IFF Spring School 2015

Jan Dhont, Gerhard Gompper, Gerd Meier, Dieter Richter,
Gerrit Vliegenthart, Reiner Zorn (Eds.)

Functional Soft Matter

This Spring School was organized
by the Institute for Advanced Simulation,
the Institute of Complex Systems,
the Jülich Centre for Neutron Science and
the Peter Grünberg Institute
of the Forschungszentrum Jülich
on 23 February – 6 March 2015.

In collaboration with
universities, research institutes and industry.

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 94

ISSN 1866-1807 ISBN 978-3-89336-999-7

Bibliographic information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Publisher: Forschungszentrum Jülich GmbH
IAS, ICS, JCNS, PGI
52425 Jülich
Phone +49 (0)2461 61-6048 · Fax +49 (0)2461 61-2410

Cover Design: Grafische Medien, Forschungszentrum Jülich GmbH

Printer: Schloemer + Partner GmbH, Düren

Copyright: Forschungszentrum Jülich 2015

Distributor: Forschungszentrum Jülich GmbH
Zentralbibliothek, Verlag
52425 Jülich
Phone +49 (0)2461 61-5368 · Fax +49 (0)2461 61-6103
e-mail: zb-publikation@fz-juelich.de
Internet: http://www.fz-juelich.de

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 94

ISSN 1866-1807
ISBN 978-3-89336-999-7

Neither this book nor any part of it may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Contents

Preface

I Introduction: Functional Soft Matter
G. Gompper, J. K. G. Dhont

A Materials

A1 Functional Nanoparticles
A. M. Schmidt

A2 Biomolecules
A. Baumann

A3 Biohybrids
T. Weidner

A4 Functional Polymers at Flat Interfaces
M. Möller

A5 Membrane Proteins & Cellular Signaling
Ch. Fahlke

B Theory

B1 Dynamics of Colloids
J. K. G. Dhont

B2 Simulation Techniques
M. Ripoll, G. A. Vliegenthart

B3 Mesoscale Hydrodynamics
R. G. Winkler

B4 Colloids and their interactions
P. R. Lang, S. De Sio

B5 Formation of Polymer membranes - Turning polymer solution into functional porosity
M. Wessling

B6 Polymerdynamics
D. Richter

C Methods

C1 Scattering
R. Zorn

C2 Light Microscopy
R. Merkel

C3 Single-Molecule Fluorescence Spectroscopy
J. Fitter

C4 Single-Molecule Mechanics and Force Spectroscopy
V. Walhorn, T. Dierks, J. Mattay, N. Sewald, D. Anselmetti

C5 Solid-State NMR
O. Petrov, M. Vogel

C6 Rheology of Complex Fluids
J. Vermant, Th. Schweizer

D Interfaces

D1 Particles at Interfaces and Membranes
T. Auth

D2 Surface Patterning
G. A. Vliegenthart

D3 Nanocomposites
G. J. Schneider

D4 Microemulsions
H. Frielinghaus

E Biomatter

E1 Active and Growing Materials
J. Elgeti

E2 Protein Folding: Kinetics, Pathways, Landscapes
J. Fitter

E3 Protein Dynamics
R. Biehl

E4 Functional Amyloids
M. Carballo-Pacheco, B. Barz, B. Strodel

E5 Protein Structure
O. H. Weiergräber, G. F. Schröder

F External Fields and Active Matter

F1 Structure and Flow
M. P. Lettinga

F2 High Pressure
G. Meier

F3 Charged Colloids and Electric Fields
K. Kang, J. K. G. Dhont

F4 Introduction to thermal gradient related effects
S. Wiegand

F5 Microgels
W. Richtering

F6 Drug delivery in blood
D. A. Fedosov, K. Müller, G. Gompper

F7 Microswimmers
G. Gompper

F8 Mechanical properties of biological protein polymers
G. H. Koenderink

G Future Technologies

G1 Self-Healing Polymers
W. Pyckhout-Hintzen

G2 Membranes for fuel cells and electrolysers
U. Reimer, W. Lehnert

G3 Polyelectrolyte membranes
O. Holderer

G4 Polymer architecture and function
E. van Ruymbeke

G5 Micro- & Nanotechnologies for Neuroscience
A. Offenhaeusser, J. Albers, S. Ullmann

Index

Preface

The spring school “Functional Soft Matter” gives an introduction to and an overview of current research topics of soft matter systems with the emphasis on biological and technological functionality. Synthetic and biological polymers, polyelectrolytes, amphiphiles and colloids are the building blocks of many materials. The understanding of their structural and dynamical properties is important for the rational design of structures with pre-described functionality and is challenging due to the enormous complexity of these systems.

The goal of this spring school is not only to teach selected topics from soft matter science and biophysics to students and postdocs in physics, chemistry and biology, but also to establish the interdisciplinary connection between these fields. This includes, in particular, to introduce biologists and chemists to physical experimental methods and theoretical modelling, and to introduce physicists to the large variety of fascinating chemical and biological phenomena.

Introductory lectures present the basics of soft matter science and biophysics. These lectures are intended to establish a common level of basic interdisciplinary knowledge. Subsequent lectures treat more advanced topics within both disciplines and emphasize interdisciplinary aspects. In addition, experimental and computer simulation techniques are introduced and explained, and examples of applications are given.

Topics of the lectures include:

- Materials
Colloidal and polymeric building blocks that are used to design functional materials have the unique property that interactions, enthalpic and entropic, can be tuned precisely, such that new structures with specific and optimized functionality can be engineered. Biomolecules like proteins, DNA and lipids are optimized by evolution and form the building blocks for complex structures with a versatile biological functionality.

- Theory & Simulations
The structure and dynamics of soft matter systems is most commonly described within the framework of classical statistical mechanics, by continuum hydrodynamics and by using scaling theory. Basic lectures present these theoretical approaches. Many phenomena, however, are too complex to obtain quantitative information from analytical theories. Therefore one has to rely on numerical techniques of which molecular dynamics-, Monte Carlo-, and mesoscale hydrodynamics-, simulations are important examples discussed here.

- Experimental Methods
The study of soft matter systems is particularly challenging since very often the microscopic building blocks are inherently complex and the relevant length and time scales in these systems span many orders in magnitude. Success requires a combined effort of preparative techniques (synthesis), the elucidation of structural and dynamical properties by scattering, microscopy, rheology and single molecule techniques.

- Interfaces
In many soft matter systems, interfaces are of major importance because the area per volume is unusually large. The school covers materials, which show a spontaneous development of interfaces such as microemulsions and cell membranes as well as situations where interfaces are created artificially in nanocomposites or nanoconfinement.

- Biomatter

Investigating the structure-property relations in biomatter is a key-tool for understanding diseases and their treatment. Therefore the school aims to provide a deeper knowledge about proteins and their properties. Advanced lectures provide information about proteins that are embedded in membranes or adsorbed on tissues. Emphasis is given on their ability to form complex 3-D structures, which rules their extraordinary dynamical behavior and their specific inter-domain dynamics. As a most prominent medical example a lecture on recent developments concerning Alzheimer's disease is presented.

- External Fields and Active Matter

Many systems, in technological applications and in living organisms, are far from equilibrium. Systems can be brought out of equilibrium by external stimuli, or can be far from equilibrium due to intrinsic properties like internal activity. Therefore, a series of lectures is provided which encompass the action of external fields on matter, to name a few: flow, pressure, electric field and temperature. In addition, intrinsically non-equilibrium systems like active swimmers and active gels will be addressed.

- Future Technologies

Functional soft matter bears a large potential for technological applications. Many of these are related to the grand challenges of the 21st century: Polymers are used as membranes in batteries and fuel cells. Bioelectronic devices are new tools for the treatment of diseases. This section also covers novel materials as polymers with tailored architecture and self-healing materials.

This school could not take place without the help and dedication of many colleagues. We would like to express our thanks to all of them for the effort and enthusiasm which they have put into the preparation and presentation of their lectures and manuscripts. Without the participation of all these colleagues, the program would not be as interesting, versatile, and attractive.

We are grateful to our colleagues from Forschungszentrum Jülich for their willingness to conduct these courses. The FZJ divisions involved in the lectures are: ICS-1/JCNS-1: Neutron Scattering, ICS-2/IAS-2: Theoretical Soft-Matter and Biophysics, ICS-3: Soft Matter, ICS-4: Cellular Biophysics, ICS-5: Molecular Biophysics, ICS-6: Structural Biochemistry, ICS-7: Biomechanics, and ICS-8: Bioelectronics and IEK-3: Electrochemical Process Engineering.
In particular, these colleagues are:

Dr. Thorsten Auth, Prof. Arnd Baumann, Dr. Ralf Biehl, Dr. Jens Elgeti, Prof. Christian Fahlke, Dr. Dmitry Fedosov, Prof. Jörg Fitter, Dr. Henrich Frielinghaus, Dr. Olaf Holderer, Dr. Kyong Kang, Dr. Peter Lang, Prof. Werner Lehnert, Prof. Pavlik Lettinga, Prof. Rudi Merkel, Prof. Andreas Offenhäuser, Dr. Wim Pyckhout-Hintzen, Dr. Marisol Ripoll, Dr. Gerald Schneider, Prof. Gunnar Schröder, Prof. Birgit Strodel, Dr. Simone Wiegand, and Prof. Roland G. Winkler.

We are very glad that several colleagues from universities and research laboratories have agreed to contribute to the program of the school:

Prof. Dario Anselmetti	Faculty of Physics, Experimental Biophysics and Applied Nanoscience, Bielefeld University
Prof. Gijsje Koenderink	Systems Biophysics Department, FOM Institute AMOLF, Amsterdam
Prof. Martin Möller	DWI Leibniz-Institut für Interaktive Materialien Aachen
Prof. Walter Richtering	Institute of Physical Chemistry, RWTH Aachen
Prof. Evelyne van Ruymbeke	Bio and Soft matter, IMCN, Université Catholique de Louvain
Prof. Annette Schmidt	Chemistry Department, Universität zu Köln

Prof. Jan Vermant	Department of Materials, ETH Zürich
Prof. Michael Vogel	Institut für Festkörperphysik, TU Darmstadt
Prof. Tobias Weidner	Surface Protein Group, MPI Polymer Research Mainz
Prof. Matthias Wessling	Aachener Verfahrenstechnik, RWTH Aachen University

We are very grateful to the board of directors of Forschungszentrum Jülich for the continuous organizational and financial support, which we have received for the realization of the IFF spring school and the production of this book of lecture notes. We acknowledge the financial support by the European Network of Excellence SoftComp. Finally, our special thanks go to Michael Beißel for the general management, and to Ursula Funk-Kath, Marie Göcking, Dorothea Henkel and Barbara Köppchen for their help in the preparation of the lecture notes.

Jan K. G. Dhont, Gerhard Gompper, Dieter Richter,
Gerhard Meier, Gerrit Vliegenthart, Reiner Zorn

January 2015

I Introduction: Functional Soft Matter

G. Gompper and J. K. G. Dhont
Institute of Complex Systems
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**
1.1 Building Blocks and Systems . . . 2
1.2 Why is Soft Matter "Soft" . . . 3
1.3 A Guide to some Recent Literature . . . 4

2 Soft Matter **4**
2.1 Dispersion Colloids . . . 4
2.2 Self-Assembling Amphiphilic Systems . . . 7
2.3 Polymers . . . 8
2.4 Soft Matter Unification . . . 10

3 Biological Matter **13**
3.1 DNA . . . 13
3.2 Proteins . . . 14
3.3 Biopolymer Filaments and Networks . . . 14
3.4 Membranes . . . 16

4 Functional Soft Matter **16**
4.1 Responsive and Adaptive Materials . . . 17
4.2 DNA-based Materials – DNA Origami . . . 18
4.3 Self-Healing Materials . . . 19
4.4 Self-Cleaning Materials . . . 20
4.5 Materials for Energy Conversion . . . 21
4.6 Biological Materials . . . 21
4.7 Active Matter – Active Materials . . . 21

5 Grand Challenges **24**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

1.1 Building Blocks and Systems

The common feature of what is referred to as "Soft Matter" is that these systems contain entities that are much larger than the typical size of simple molecules. There are a variety of different types of soft matter systems, depending on the type of this "large entity".

Probably the simplest example are the classical *colloidal systems* consisting of small solid, spherical and rod-like particles (the "colloidal particles") of inorganic material, such as gold particles (about $1 - 10\ nm$), silica particles (about $10 - 1000\ nm$ in radius), which can be considered as small "sand particles", and aluminum-oxide rods (with a length of about $50 - 500\ nm$).

The second large class of Soft Matter systems are *polymeric systems*. These macromolecules consist of long chains of identical monomers, which are covalently linked by a carbon backbone. The large variety of polymer materials arises in part because of many possible types of side groups. Many polymers are linear, but there are also branched polymers, block copolymers consisting polymers of different chemical structure linked together, star polymers, etc.

Many Soft Matter systems are *self-assembled systems*. The classical example is mixtures of surfactant molecules with oil and/or water. Surfactant molecules assemble at the oil-water interface, because they consist of a hydrophobic and a hydrophilic part. Self-assembly of *amphiphilic* molecules can lead to large structures like spherical micelles, an aggregate of surfactant molecules with a diameter of a few nm's, and microemulsion droplets, consisting of an oil droplet with a surfactant monolayer at its surface with a diameter of up to a few ten's of nm's.

Another common feature of soft matter systems is that the macromolecules and aggregates are embedded in a solvent – most of the time water. These soft matter systems are therefore also referred to as "complex fluids". This implies that the dynamical behavior is often strongly affected by the solvent hydrodynamics. There are exceptions, like polymer melts, which do not contain solvent but are still fall within the class of soft matter.

There is an abundance of macromolecules that play an essential role in *living matter*. DNA is a helical macromolecule that consists of paired nucleic acids, and carries the genetic information. Depending on genetics, many different types of proteins are synthesized within the cell, which are necessary for the functioning of the cell. Proteins themselves are macromolecules consisting of a linear covalent attachment of about 20 types of different amino acids. The order of these amino acids defines the protein. The linear amino-acid chain folds in a way that makes the proteins functional for its special task within the cell. Proteins can also assembly in larger fibrillar structures, and form networks that are necessary for directed transport by molecular motors and which are important for cell motility and cell division. Amphiphilic molecules (lipids) assembly to membranes, which form the outer part of a cell, as well as membranes within the cell for various purposes. Within cellular membranes, various proteins reside, which, in cooperation with proteins in solution, regulate for example the transport of ions through the outer membrane of a cell. All of these macromolecular "living systems" constitute very complex soft-matter systems, with very complicated, specific inter-macromolecular interactions.

There are nowadays several directions in which soft-matter science is developing:

(i) Due to the macromolecular nature of living matter, part of soft-matter science is developing towards biophysics, where complex biomimetic and biological systems are studied.

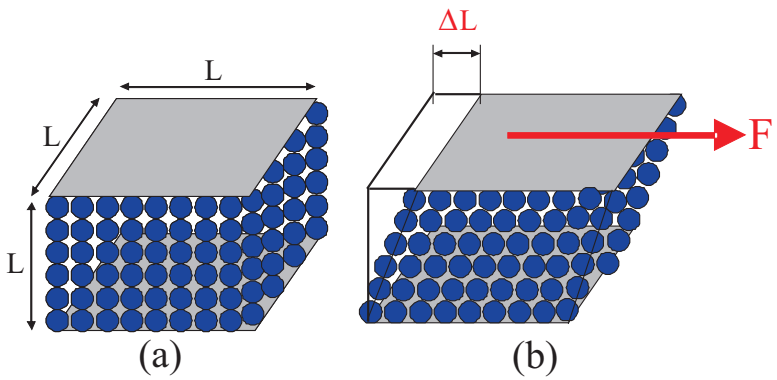


Fig. 1: *(a) A cube of crystalline material. (b) Deformation of a crystal by a shearing force F.*

(ii) The behavior of mixtures consisting of several different types of soft-matter systems (like colloids and polymers, or colloids and liquid crystalline materials) is very rich due to the complex interactions and the many parameters that can be tuned. This offers the possibility to obtain many new material properties.

(iii) Soft-matter systems respond quite sensitively to external fields, and are thus easily taken far away from equilibrium. The response to external fields and non-equilibrium phenomena can thus be studied reletively easily, but also be exploited in structure formation, for example in the directed self-assembly of nano-structured materials

(iv) By a combination of such properties, functional Soft-Matter systems and smart materials can be constructed, which includes responsive and adaptive materials, self-healing and self-cleaning materials, and materials for energy applications. Composite smart materials are therefore interesting with many possible technological applications.

1.2 Why is Soft Matter "Soft"

Why are these systems that contain large entities called "soft"? The reason for this is as follows. Let us compare the energy that is needed to deform the surface of a common molecular crystal (like sodium chloride) and of a crystal of giant spherical "molecules" with a size of, say, 1 micron. The size of ordinary molecules is typically one Angstrom, which is a factor of 10^4 smaller than the colloidal spheres. Suppose that a cube with faces of length L is deformed by means of a small external shearing force F, as depicted in Fig. 1. The shear modulus μ of the crystal relates the displacement ΔL to the force as,

$$\frac{F}{L^2} = \mu \frac{\Delta L}{L} \tag{1}$$

Thus, the shear modulus has the dimension energy/(length)^3. The energy scale is here the typical interaction energy of neighboring molecules, which is on the order of the thermal energy $k_B T = (1/40)eV$ for not too low temperatures. The length scale is proportional to the lattice constant, since this is the only characteristic length scale in a crystal. This simple estimate indicates that the shear modulus of a colloidal crystal is about 12 (!) orders of magnitude

smaller than that of a usual crystal. This result indeed reproduces experimentally observed values quite well. Therefore, colloidal crystals are much more easier to deform, that is, they are extremely soft and can be destroyed by mechanical means very easily. This is the consequence of the fact that the colloids are so much larger as compared to simple molecules.

1.3 A Guide to some Recent Literature

Several books, which cover different areas of soft matter and cellular biophysics, have been published in recent years. A (necessarily incomplete) list is:

- Colloids:
  Russel (1987,1989) [1,2], van de Ven (1989) [3], Hunter (1989) [4], and Dhont (1996) [5].

- Polymers:
  Flory (1969) [6], de Gennes (1979) [7], Ferry (1980) [8], Doi & Edwards (1986) [9], de Cloizeaux & Jannink (1987) [10], Grosberg & Khokhlov (1997) [11], and Hamley (2005) [12].

- Amphiphilic Systems and Membranes:
  Meunier, Langevin & Boccara (1987) [13], Gelbart, Ben-Shaul & Roux (1994) [14], Gompper & Schick (1994) [15], Safran (1994) [16], Lipowsky & Sackmann (1995) [17], Kumar & Mittal (1999) [18], and Nelson, Piran & Weinberg (2004) [19].

- General Soft Matter:
  Elias (1990) [20], Israelachvili (1992) [21], Evans & Wennerström (1994) [22], Chaikin & Lubensky (1995) [23], Cates & Evans (2000) [24], Dhont, Gompper & Richter (2002) [25], Gompper, Kaupp, Dhont, Richter & Winkler (2004) [26], Witten & Pincus (2006) [27], Poon & Andelman (2006) [28], Hamley (2007) [29], Dhont, Gompper, Nägele, Richter & Winkler [30], Dhont, Gompper, Lang, Richter, Ripoll, Willbold & Zorn [31], and series of books on soft matter edited by Gompper & Schick since 2005 [32–35].

- Cell Biophysics:
  Howard (2001) [36], Boal (2002) [37], Schiessel (2003) [38], Phillips, Kondev, Theriot & Orme (2008) [39], Sackmann & Merkel (2010) [40], Schiessel (2013) [41], and Nelson (2015) [42].

2 Soft Matter

The "classical" *soft matter* systems, which have already been studied since the beginning of the 20th century are dispersion colloids, amphiphilic systems and polymers. We will briefly introduce these systems here.

2.1 Dispersion Colloids

As mentioned above, the size of a colloidal particle is by definition much larger than the size of solvent molecules (about $0.1\ nm$). An upper bound for the size of colloidal particles is set by the requirement that they exhibit vivid thermal motion. This thermal motion is also referred to as "Brownian motion" , named after the English Botanist who observed this motion for pollen

grains in 1827. Thermal motion of colloidal particles becomes weaker as they become larger, since the forces that solvent molecules exert on the surface of the colloidal particle tend to average out. This limits the largest size of colloidal particles to about $10 - 20$ micron. Hence, colloidal systems can be regarded as solutions of very large, thermally active molecules. Such solutions are commonly referred to as *dispersions* or *suspensions*. Colloidal systems thus exhibit phase transitions which can be described on the basis of thermodynamics and statistical physics. Phase transitions (like crystallization, gas-liquid phase separation and formation of nematics and smectics) and critical phenomena (like critical slowing down of diffusion and critical opalescence) are amongst the phenomena that colloidal systems have in common with molecular systems. Colloids can thus be used to study fundamental questions which are also relevant for molecular systems. Due to their large size and slow dynamics, the experimental study of colloids is usually much simpler than for molecular systems (for example, instead of X-ray or neutron scattering, light- scattering and microscopy can be used in many cases, and times that are involved are in the second range, instead of the microsecond or even picosecond range).

The lower limit of approximately $1\ nm$ on the size of a colloidal particle assures that the fluid in which the colloidal particles are embedded can be treated as a structureless continuum. Since the typical size of a solvent molecule is 1 Angstrom, the colloidal particle can be regarded as a macroscopic object on the solvent length scale (an exception on this would be a solvent that itself is close to its critical point). Interactions between colloidal particles and solvent molecules can thus be treated on the basis of the Navier-Stokes equation for the solvent. Moreover, relaxation times of degrees of freedom of the solvent molecules are much smaller than those for the colloidal particles. Dynamics and non-equilibrium phenomena can thus be described by equations of motion in which the degrees of freedom of the solvent molecules are integrated out. Both for the structure and dynamics, a colloidal system of identical colloidal particles can thus be regarded as an effective one-component system. The solvent just acts as a "thermal motor" that drives the motion of the colloidal particles.

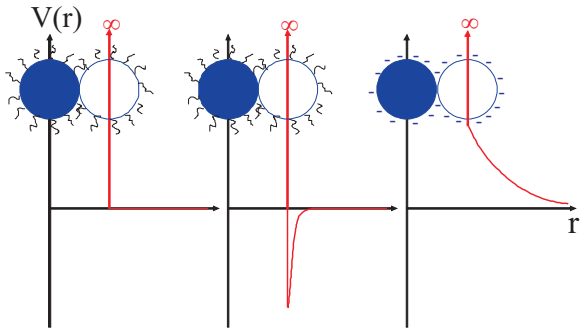


Fig. 2: *Three possible pair-interaction potentials $V(r)$, where r is the distance between the centers of two spherical colloids. (a) A hard-core interaction potential, where the core of the colloidal sphere is coated with short polymers chains and the solvent is a good solvent for the coat polymer. (b) A short-ranged attractive potential on top of the hard-core potential, which is achieved when the core is coated with polymers and the solvent is a bad solvent for the coat polymer. (c) Like-charged colloids repel each other, and the range of the repulsive potential can be varied through variation of the salt concentration.*

The interest in colloidal systems goes well beyond the above mentioned possibility to gain insight in fundamental questions that also apply to molecular systems. Colloidal systems have their own specific properties that have no counterpart in molecular systems. There are two reasons why colloids exhibit specific phenomena not found in molecular systems : (i) the various kinds of interaction potentials that can exist between the colloidal particles which do not exist for molecules, and (ii) the slow dynamics of the colloidal particles, giving rise to non-equilibrium phenomena that are not found in molecular systems.

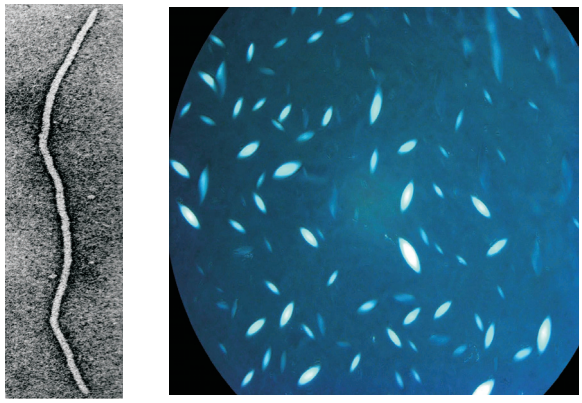


Fig. 3: *(a) A fd-virus particle has the shape of a long, semi-flexible rod. It has a diameter of about* $6\ nm$*, a length of* $880\ nm$*, and persistence length of about* $2.2\ \mu m$*. (b) Nematic "tactoids" (the bright regions) of fd-virus particles floating in an isotropic fd-virus solution (the blue region). The tactoids are about* $50\ \mu m$ *long. Picture courtesy of K. Kang (FZ Jülich).*

The interaction potential between molecules is set by their electronic structure. The interaction potential between colloidal particles, however, can almost be chosen at will, ranging from purely long-ranged repulsive to very short-ranged attractive. The potential can be tuned in a number of ways. First of all, the surface of a colloidal particle can be covered by short or long polymer chains or charged groups. When the solvent is a good solvent for the polymers attached to the surface, and the polymer chains are very short in comparison to the size of the core of the particle, the interaction potential is essentially a so-called hard-core potential: the interaction potential is infinite when the inter-particle separation is smaller than the size of the core and zero otherwise (see Fig. 2a). When, on the other hand the solvent is a bad solvent for the polymer, the interaction potential has a very short-ranged attractive part, on top of the hard-core repulsion (see Fig. 2b). Such a systems exhibits gas-liquid phase separation and it can form a gel where the colloid particles "freeze" into a long-lived, non-equilibrium state. The existence of a gel-state is an example of a state which is not found for the usual molecular systems.
In the good solvent case, one can experimentally study nucleation and crystal growth kinetics for a model hard-core system, while for the poor solvent case one can study gas-liquid phase separation kinetics and gel formation. Secondly, one can change the potential by external means. For example, the quality of the solvent for the polymer attached onto the surface of a colloidal particle can be changed by changing temperature: the depth of the attractive part in Fig. 2b the potential is then changed. One could also change the refractive index of the solvent,

thereby changing the strength of the always present van der Waals attractions between the cores of the colloidal particles. When the surface of the colloidal particles is charged, one can change the range of the repulsion by adding or removing salt from the solvent (see Fig. 2c). Another possible way to manipulate both the range and strength of a predominantly attractive potential is to add free polymer to the solvent. So-called "depletion attractions" between the colloidal particles are now induced. The range of attraction is set by the size of the polymers while the interaction strength is set by the number concentration of polymers. In this way one can chose an interaction potential depending on the kind of phenomena one wishes to study, or find new kinds of phenomena on changing the potential. The above not only holds for spherical colloidal particles, but also for non-spherical particles. Non-spherical colloidal particles that are commonly used are bio-polymers (like fd-virus (see Fig. 3a), tobacco mosaic virus (TMV) and DNA fragments) or inorganic particles. Rigid rod-like colloids exhibit an isotropic-nematic phase transition. In Fig. 3b an example of coexistence of droplets containing a nematic of fd-virus particles and an isotropic phase is shown.
The very slow dynamics of colloids gives rise to non-equilibrium phenomena that are not found in molecular systems. The micro-structural arrangement of colloidal particles relaxes typically in a time range of seconds. This results, for example, in a non-linear response of colloidal systems to a shear-field. For molecular systems such a non-linear response would occur only at unrealistically high shear-rates and frequencies. A practical application of this kind of non-linear behavior of colloids is paint. At high shear rates the viscosity of paints should be small, so that one can easily bring paint onto the wall; after that, the paint should stay on the wall and not simply slide off the wall under gravity. Hence, at low shear rates the viscosity of paint should be large. Such complex fluids are called "shear-thinning".

2.2 Self-Assembling Amphiphilic Systems

Amphiphilic molecules consist of a hydrophilic , "water-loving" part and a hydrophobic "water-hating" part, see Fig. 4. The hydrophobic part typically consists of one or two hydrocarbon chains. The hydrophilic part is a polar group in the case of non-ionic amphiphiles, and a ionizable group, which dissociates in water, in the case of ionic amphiphiles — such as the molecule shown in Fig. 4. Amphiphilic molecules therefore prefer to sit at the interface between hydrophilic and hydrophobic solvents. Their name derives from the Greek word "amphiphile", which means "loving both". These molecules are also often called "surfactants", for *surf*ace *act*ive substance, since they assemble at surfaces and reduce the surface tension.

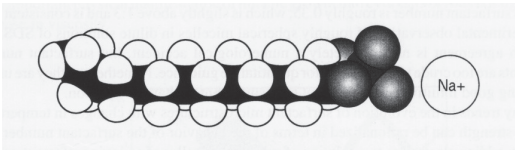


Fig. 4: *Sodium dodecyl sulfonate (SDS) is an typical example of an ionic amphiphile. The dissociable headgroup (right) is linked to one hydrocarbon chain (left).*

Structure formation in amphiphilic systems occurs by supramolecular aggregation. The dominant driving force is here the *hydrophobic effect*, which makes mixtures of polar and apolar molecules highly unfavorable. The underlying physical mechanism is not very well understood,

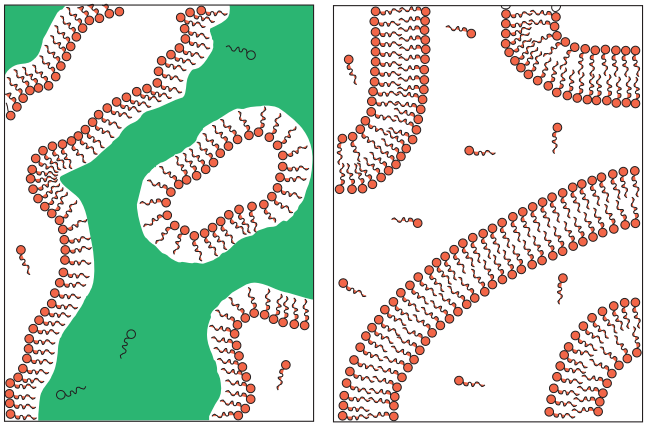


Fig. 5: *Amphiphilic molecules self-assemble into complex aggregates due to the hydrophobic effect. In mixtures with oil and water, they form monolayers at the water-oil interface. In mixtures with water, they form bilayers.*

but is most likely due to the perturbation of the network of hydrogen bonds in polar solvents like water by an apolar solute particle. This effect is responsible for oil and water not to be miscible at ambient temperatures. Amphiphilic molecules preferentially adsorb at the oil-water interface, because in this way both parts of the molecule can be in their favorable environments. The adsorption of surfactants reduces the tension of the oil-water interface, because at the location of the amphiphilic molecule, the unfavorable direct contact between oil and water can be avoided. The interfacial tension decreases with increasing surfactant concentration, until a state is reached in which the amphiphilic molecules are so densely packed that no space is available for more molecules to squeeze in (compare Fig. 5). This state has been called by Schulman a *saturated monolayer* [43]. The interfacial tension is extremely low in this case, or vanishes completely, so that new phases become possible, in which oil and water are mixed on a mesoscopic length scale, typically $1 - 100\ nm$. Such phases are called *lyotropic phases* or *microemulsions*. A schematic phase diagram is shown in Fig. 6. As can be seen, a whole variety of structures can be formed, ranging from spherical entities to cylinders, sheets and bi-continuous structures. In binary mixtures of water and amphiphile, no oil-water interface exists for the surfactants to assemble. Therefore, they form spherical aggregates in this case, called *micelles*, or bilayers (compare Fig. 5). Thus the amphiphilic molecules form in this case a favorable environment for themselves, by sticking their hydrocarbon tails together and exposing only the polar headgroups to the water. Amphiphilic bilayers are the basic building block of all biological membranes, as will be discussed in more detail in Sec. 3 below.

2.3 Polymers

In the simplest case, polymers are long, linear chain molecules, which are built from a repeat unit, called a monomer. Such polymers are *linear homopolymers*. Many properties of polymers can be custom-tailored by the use of different monomers in the polymerization process, or by a variation of the polymer architecture which can be linear, star-shaped, H-shaped, like a bottle-brush, etc.

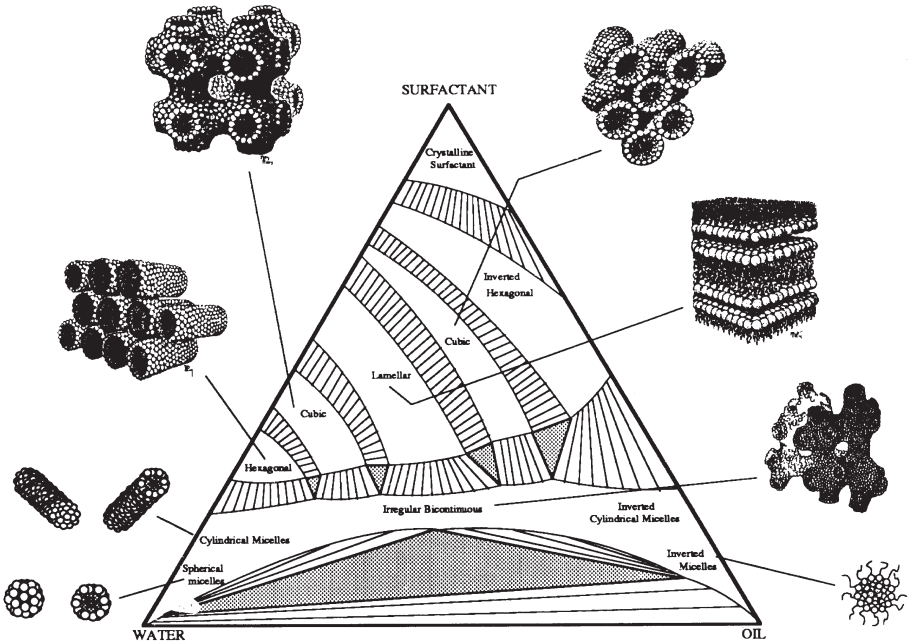


Fig. 6: *Schematic phase diagram of ternary amphiphilic systems. From Ref. [44].*

The individual monomers or building blocks are bound together by covalent bonds. This usually does not restrict rotations around the bond axis, so that such rotations are still possible. Therefore, long chains have a very large number of internal degrees of freedom, which lead to an enormous number of spatial conformations of all the monomers. The entropy of such a chain is directly related to the vast number of conformations, and is therefore also large. The free energy of a chain molecule is therefore not only determined by the internal interaction potentials, but to the same extent by the chain entropy. On distances, which are large compared to the size of a monomer, the chemical structure of the building block plays only a minor role, and the properties of the chain are mainly determined by the statistical mechanics of the chain, i.e. essentially by the chain entropy.

The *central limit theorem*, also called the law of large numbers, states that the most probable spatial conformation of a large number of monomers is governed by Gaussian statistics. When external forces act on such a chain, it is stretched out and thereby looses part of its entropy, which thus favors a return to the most probable, coiled-chain conformation. This is equivalent to an entropic force, which is responsible for rubber elasticity.

Many properties of polymers are a direct consequence of the law of large numbers and the laws of statistics, and are therefore *universal valid*. For a polymer chain, which is in a coil conformation (compare Fig. 7), the radius of gyration R_g or the average end-to-end distance R_e depend, for example, on the number N of monomers in the form of a power law,

$$R_e \sim N^{\nu} . \tag{2}$$

For Gaussian chain statistics, the exponent is $\nu = 1/2$. The entropic nature of rubber elasticity

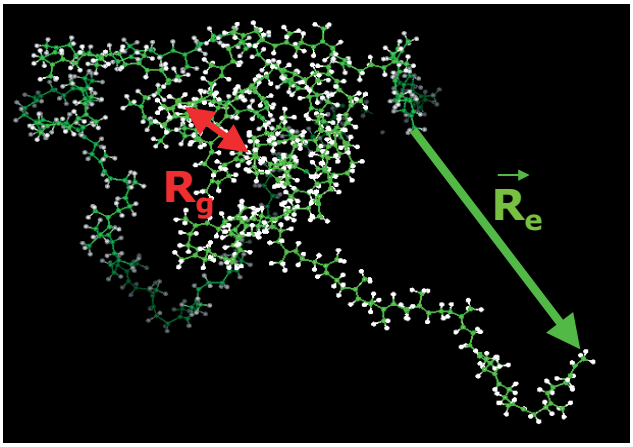


Fig. 7: *The typical conformations of long, flexible polymers (such as polyisoprene shown here) are random coils. These conformations can be characterized by an end-to-end distance R_e and a radius of gyration R_g.*

implies that the restoring force is inversely proportional to the mesh size of the polymer network. As a consequence of the chain topology, the viscosity of a polymer melt of linear chains again depends on the chain length in the form of a power law. There are many other examples of such universalities, which are independent of the type of polymer.

In addition, properties which derive from the specific chemical structure of the monomers of course also play a role. The microstructure of a polymer, which can be regular or random, determines for example whether the polymer crystallizes or not. The miscibility of polymer systems is extremely sensitive to the chemical structure of the chain. Also, transport properties like diffusion are determined by a friction coefficient, which is specific for a given material. This interplay between universal and specific properties is one of the major challenges of polymer science. The goal is to employ both of these aspects to construct polymer materials which just the desired properties.

Plastics and polymers play a very important role in our daily lives. Polymer materials are used everywhere, because they are durable, simple to produce and easy to process, and because their mechanical properties are very versatile and can be varied over a wide range. The same polymer can, for example, flow viscously, react like a rubber, or show a pronounced brittleness. Polymeric fibers can have the strength of steel, polymeric electrolytes have an enormous absorption strength, which is employed for example in baby diapers, polymeric additives make diesel fuel winter hard, polymers are used as viscosity modifiers in motor oil, polymers serve as construction material for a large variety of applications, and polymers are suitable as template materials in nano-technology due to their self-assembling properties.

2.4 Soft Matter Unification

As already mentioned in subsection 1.1, in the last 10-20 years, the previously largely independent research fields of colloids, membranes, microemulsions and polymers have been integrated in the new research field soft matter. It has been realized that many phenomena in these sys-

tems have the same underlying physical mechanisms, so that similar effects do not have to be discovered and understood anew in each subfield. May be even more importantly, it has been recognized that combinations of these classical systems, like for example polymers and colloids, exhibit new properties which are not found in each of the systems separately.

Two aspects are particularly important:

Macromolecular Toolbox — For many of the recently studied systems is it very difficult to classify them into one of the classical categories. As rod-like colloids become longer and longer, their flexibility becomes increasingly important. As soon as their length exceeds the persistence length, they behave as a semi-flexible polymer, and even longer ones as flexible polymers. As an example, consider the fd-virus of Fig. 3. Is it more appropriate to call it a colloidal particle or a semi-flexible polymer? Similarly, amphiphilic properties can be varied gradually. Amphiphilic molecules can be short, such as the classical surfactants, or very long, such as amphiphilic block copolymers. Also, it is nowadays possible to synthesize so-called spherical colloidal Janus particles, where the two faces of a colloidal sphere can be made hydrophilic and hydrophobic (in Roman mythology Janus is the god of change and transition. He is often portrayed with two faces looking in opposite directions). These examples show that the macromolecular construction kit has grown beyond its original classification, and that there is now a large variety of molecules and building blocks available, as shown schematically in the "magic triangle" in Fig. 8.
As another example consider F-actin and worm-like micelles. F-actin (filamentous actin) is a very long and stiff object. It can *not* be classified as a classical colloidal rod-like particle, however, since it changes its length through polymerization. The polymerization process is at the origin of the force that networks of F-actin exert on cell membranes which in turn leads to motion of the cell. Similarly, worm-like micelles can break and reform, which is especially important when they are subjected to shear flow. The flow behavior of these system is to a large extent determined by scission and recombination kinetics of these elongated surfactant aggregates. The polymerization process and scission and recombination of worms are additional complications that are absent in classical systems of colloidal rods.

Universality — Many phenomena of soft matter are *independent* of the chemical structure of the molecular building blocks. A simple example are polymers. As discussed in Sec. 2.3, when the chains are long enough, the average end-to-end distance as well as other polymer properties are found to depend on the monomer number by a power law, with an the exponent ν, which has the same value for *all* polymers of the same topology and the same class of interactions, independent of the chemical architecture of the monomers. Similar relations can be found for membranes and colloids, where also many effects are independent of the chemistry of surfactants or colloidal particles.
The reason for the universality is again the large characteristic length scale of soft matter systems, which implies that a very large number of atoms or atom clusters have to cooperate to make things happen. In the simple case of a linear polymer chain without self-avoidance, the *law of large numbers* implies $\nu = 1/2$. When self-avoidance is taken into account, cooperativity effects become even stronger, and the exponent increases to $\nu = 0.6$. Therefore, from a theoretical point of view, soft matter physics is the domain of *Statistical Mechanics*, which provides the framework in which cooperative phenomena can be understood.
As already mentioned above, there is also a universality among the different classes of materials. As an example, consider a mixture of two homopolymers A and B (i.e. polymers made of a single type of monomer) with a so-called A-B diblock copolymer, in which two different

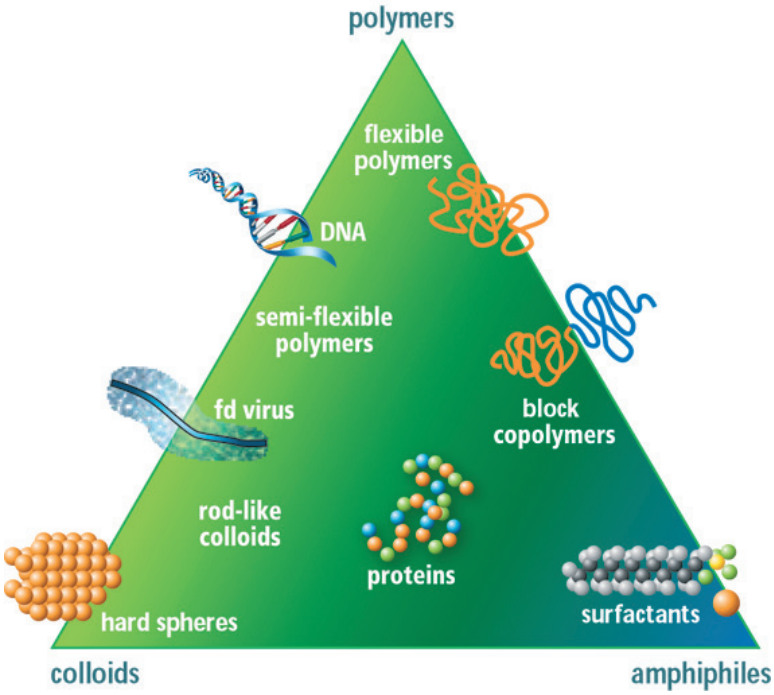


Fig. 8: *The "magic triangle" of soft matter, which shows that the classical fields of colloids, polymers and amphiphilic systems have merged into one. The aspect ratio (length/width) increases in the vertical direction, the amphiphilicity in the horizontal direction. From Ref. [45].*

homopolymers are linked together chemically to form a single, correspondingly longer chain. If the two A- and B-polymers are immiscible, which is often the case, then a mixture can be achieved by addition of small amounts of diblock copolymers. Thus, block copolymers play a very similar role in this case as amphiphilic molecules do in mixtures with oil and water. A second example is the viscosity of dilute polymer solutions. The dependence of the viscosity on the polymer concentration is exactly the same as in a dilute colloidal suspension. Thus, in the dilute regime, polymers behave like hard spheres. Finally, the worm-like micelles mentioned above, which can be so long that they exceed their persistence length many times and then behave as "living polymers", i.e. in many respects they behave as polymers, but they have the additional dynamical property of fission of long micelles into shorter ones, and vice versa the fusion of short micelles into longer ones.

3 Biological Matter

Macromolecules are the building blocks of biological cells. DNA is a linear polymer, which in its sequence of nucleic acids stores the genetic information. There is a complex machinery in the cell which transforms this information into amino-acid sequences, which constitute functional units – the proteins. There is large variety of proteins in the cell, each of which performs a specific task. Some proteins assemble into long filaments, like actin filaments and microtubules. Another type of self-assembly occurs for lipid molecules, which due to their amphiphilic character form bilayer membranes. These membranes form the outer envelope of the cell – the plasma membrane – and are present in the interior of the cell to form compartments. A special class of proteins are membrane proteins, which play an important role to regulate transport through the membrane.

3.1 DNA

The most well-known and important polymer in living systems is certainly the carrier of the genetic information, *desoxyribonucleic acid (DNA)*, see Fig. 9. It consists of a sequence of four monomers, the nucleotides adenine, thymine, guanine and cytosine, which in pairs form the famous double helix. Three subsequent bases encode for one amino acid. The genetic code specifies 20 standard amino acids.

In eukaryotic cells, DNA is tightly packed in the form of chromosomes within the nucleus [38, 41]: the linear length of stretched DNA within a chromosome is typically of the order of centimeters, and is contained within the volume of the nucleus which has a diameter of a few microns. Such a tight packing is possible through the interactions of DNA with so-called chromatin-proteins. It is still a matter of debate how the mechanism of tight packing functions. In prokaryotic cells, DNA is present as circular molecules, which are super-coiled through interactions with enzymes. These helical super-coiled DNA's float freely in the cell's plasma. It is thus clear that DNA only functions with the help of complex interactions with several types of proteins. This is the case for almost all processes in living matter: many different types of molecules work together to maintain life.

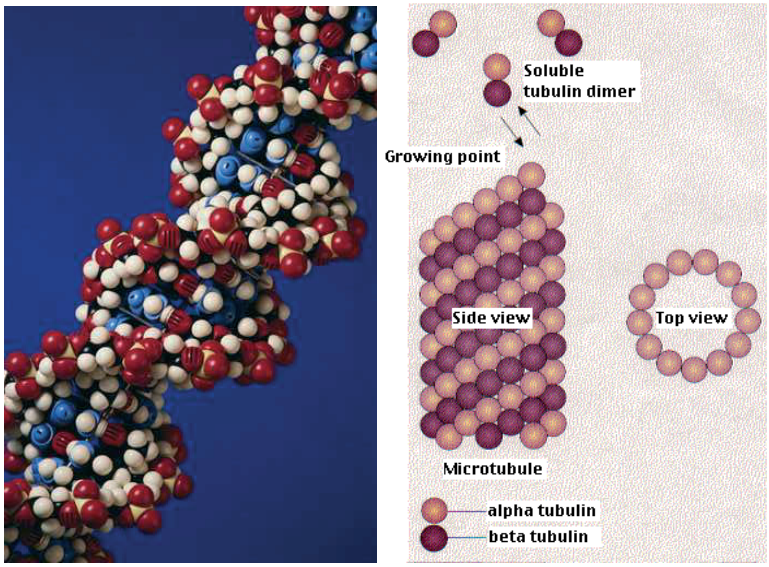


Fig. 9: *Structure of biopolymers. Left: DNA is a double-stranded helix. Right: Microtubules are assembled from dimers of α and β tubulin;* 13 *protofilaments from a hollow tube.*

3.2 Proteins

Proteins were first described by the Dutch chemist Gerhardus Johannes Mulder and named by the Swedish chemist Jöns Jakob Berzelius in 1838. Proteins (also known as polypeptides) are linear chains of amino acids, which fold into a globular or fibrous form. The amino acids are covalently connected by peptide bonds. Proteins are formed from a "library" of 20 standard amino acids.

The content and sequential order of amino acids define the "primary structure" of a protein. One of the most distinguishing features of polypeptides is their ability to fold into a globular state. This folding occurs on two hierarchical levels. The "secondary structure" consists of α-helices, pieces of the chain which form tight staircase-like cylindrical units, and β-sheets, which form ribbon-like units. One of the important forces that stabilizes the secondary structure are hydrogen bonds between the amino acids. These elements then order into a three-dimensional arrangement, which defines the "tertiary structure". The function of a protein strongly relies on the correct tertiary structure. Two examples of proteins with high content of α-helices and β-sheets are shown in Fig. 10.

The extent to which proteins fold into a defined structure varies widely. Some proteins fold into a highly rigid structure with small fluctuations and are therefore considered to be single structure. Other proteins undergo large rearrangements from one conformation to another. Not all proteins require a folding process in order to function.

3.3 Biopolymer Filaments and Networks

An important biopolymer is *actin*, one of the three major components of the cytoskeleton; it participates in many important cellular functions, including muscle contraction, cell motility, cell division and cytokinesis, vesicle and organelle movement, cell signaling, and the establishment



Fig. 10: *Aquaporin 1 is a member of the Aquaporin family. It was first discovered in red blood cells. In 2003, Peter Agre was awarded the Nobel prize for discovering this "water channel". This channel is used for rapid transportation of water across the cell membrane. It is very specific and only allows passage of water molecules, but is impermeable to hydrogen ions. Note the large content of α-helices. From Ref. [46].*

and maintenance of cell junctions and cell shape.

A second component of the cytoskeleton are *microtubules*, which are responsible for the transport of vesicles within the cell; also, during cell division, microtubules are involved in the process of the separation of the chromosomes. The structure of microtubules consists of a hollow tube of dimers of the protein subunits α and β tubulin, which are arranged in protofilaments, see Fig. 9.

An important aspect of biopolymers are their mechanical properties, which can be characterized by the persistence length, *i.e.* the length below which a polymer behaves essentially like a stiff rod, whereas it behaves like a random coil on larger scales. The persistence lengths of DNA, actin and microtubules are $50\ nm$, $10-20\ \mu m$ and about $1\ mm$, respectively. These numbers can be connected to their functions and location in the cell. Very long DNA strands have to be curled up in the cell nucleus, which is only a few μm in diameter (the total length of a single DNA strand in the human genome is $1.8\ m$); DNA therefore has to very flexible. On the other hand, microtubules have to very stiff, because they are the "highways" for vesicle transport; their persistence length therefore exceeds the cell diameter.

In contrast to synthetic polymers, biopolymers are often not permanently, "passively" polymerized. Instead, they are dynamic states of an active polymerization process, which proceeds with different rates at the two ends of the *polar* filaments. Both actin and microtubules belong to this class of "tread-milling" filaments.

3.4 Membranes

The main component of a cell membrane are lipids, amphiphilic molecules with two hydrocarbon tails (see Fig. 11), which form bilayers in water. This was first shown experimentally by Gorter and Grendel (1925). They first extracted lipid molecules from the plasma membrane of red blood cells. By pouring the lipids onto a water surface, and comparing the area of the resulting patch with the surface area of the original cell surfaces, they concluded that the cell membrane consists of a lipid bilayer. Later it turned out that their arguments contained several errors, they had for example underestimated the size of the original cell surface. However, these errors canceled out each other to a large extent, so that their conclusion was nevertheless correct. Nowadays, it is well established that the major component of the cell membrane consists of a lipid bilayer. Within cell membranes there are usually a variety of other molecules embedded (an artistic impression is given in Fig. 12). The macromolecules and macromolecular complexes that are embedded in the membrane regulate exchange of matter with its surroundings. For example, ion channels are complexes which regulate exchange of specific ions.

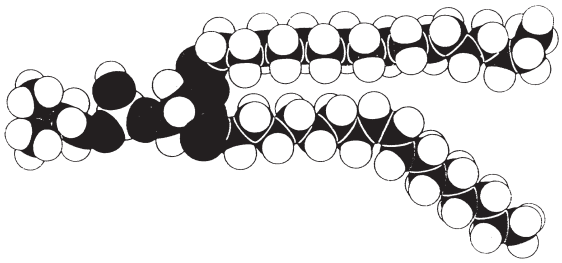


Fig. 11: *Phosphatidyl-choline (PC) is a typical example of a membrane lipid. The polar head (left) is connected to two hydrocarbon tails (right).*

For the same reason as sketched above for colloidal crystals, membranes consisting of (bio-) polymers are quite easily deformable. A typical example of a such a biological membrane is the "skin" of a red blood-cell, see Fig. 13, the size of which is in the few micrometer range. Due to the flexibility of the membrane, a red blood cell can penetrate through openings with a size four times smaller than its own size. In addition, they are easily deformed by solvent flow, which is important for blood flow in microvessels. An example of blood-cell deformation in flow is shown in Fig. 13. The figure also shows the flow-induced interaction of red blood cells with other particles in the blood stream, for example with drug carriers or platelets.
Biological membranes have a very complicated structure. Fundamental insight in properties of membranes must thus be obtained through the study of more simple "physical" membranes.

4 Functional Soft Matter

Since the times when W. H. Carothers produced the first nylon polymers (1931) and K. Ziegler and G. Natta discovered catalyser systems, which polymerize polyethylene at room temperature and ambient pressure (1953) – opening the way for the industrial production of commodity

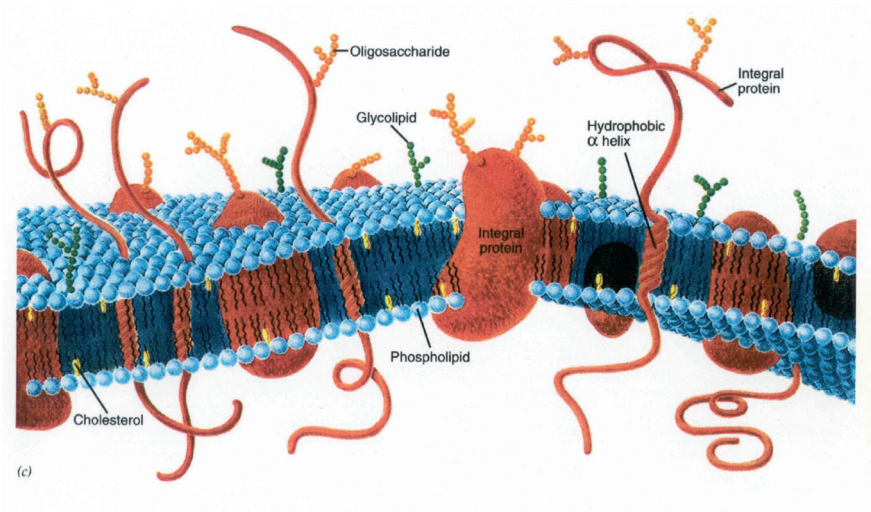


Fig. 12: *Artistic impression of a cell membrane in which many macromolecules and macromolecular complexes are embedded. From Ref. [47].*

polymers –, the amount of polymers produced per year has grown very rapidly and now reaches about 1 billion (10^9) tons. Of course, all the polymer materials produced are "functional" in the sense that they are useful in our daily lives – like plastic bags, back-packs, rain gear, blankets, plastic bowls, and auto parts.

However, when we are talking about "functional soft matter", or somewhat synonymously "smart materials", then we mean much more advanced materials. Such functional matter includes, for example, responsive and adaptive materials, self-healing and self-cleaning materials, and materials for energy conversion.

4.1 Responsive and Adaptive Materials

There is an abundance of applications of different kinds of soft functional materials based on their response to changing environmental conditions ("adaptive materials") and to external stimuli ("responsive materials"). These materials find their applications in drug delivery, tissue engineering, diagnostics, bio-electronics, optical systems, biosensors, mechanical systems, coatings and textiles.

Colloidal and polymeric materials can, for example, regulate transport of ions and molecules, they can change binding conditions and adhesion of various molecular species depending on environmental conditions, and they can generate optical and electrical signals due to chemical stimuli. As an example of changing binding conditions, Pickering emulsification (i.e. emulsification with colloidal particles instead of surfactants) can stabilize oil-in-water droplets, which can be easily de-stabilized by additives. In such a way, crude oil can be separated in oil and a water phase, the latter of which contains the Pickering nano-particles that can subsequently be re-used. An example of response to an external field can be found in the fluids of shock absorbers, where magnetic beads form strings under the action of a magnetic field, which enhanced the viscosity of the suspension, and hence serves as a strong increase in friction

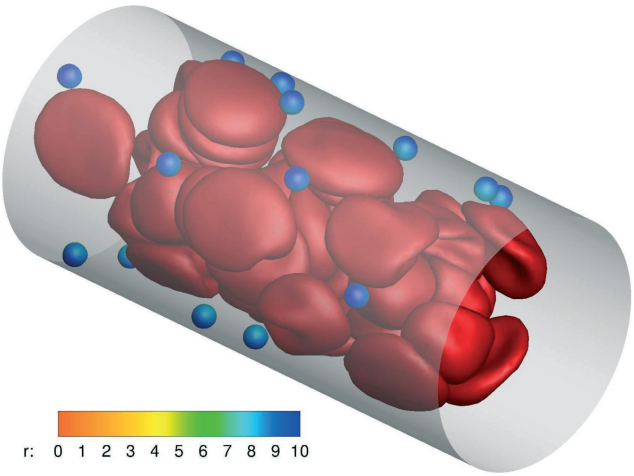


Fig. 13: *Margination of Drug Carriers in Microchannel Flows. Red blood cells (erythrocytes) have a characteristic, biconcave disk shape (discocyte) at rest, and are strongly deformed by flow. Carriers interact with the red blood cells and can thereby be pushed toward the wall. The carriers are colored according to their radial position r, as indicated, which demonstrates their strong margination. From Ref. [48].*

(magneto-rheological fluids). Another example concerns targeted drug delivery with a high spatial and temporal resolution, which can be achieved with light-responsive materials, which either degrade or dissociate upon irradiation.

4.2 DNA-based Materials – DNA Origami

DNA can be used to construct a large variety of macromolecules with complex architecture, due to the simple "click chemistry" (or "Watson-Crick binding") involved in matching of DNA base-pairs [49]. The DNA-based click chemistry opens the way to engineer macromolecular shapes that would not be feasible with purely synthetic materials. These DNA-based macromolecules are commonly referred to DNA-origami. The simplest of such macromolecules are, for example, two rods linked together with a flexible spacer, where the two rod-like ends consist of double-stranded DNA, while the middle part is unmatched, or a Y-shaped geometry based on the same principle with either a flexible part between the three legs (see the upper panel in Fig. 14(a)), or a stiff connection leading to a planar structure (see Fig. 14(b)). Much more complex origamis based on base-pair click chemistry can be constructed. The folding of a very long backbone DNA strand enables the construction of large, two-dimensional scaffolds [50], an example of which is given in Fig. 14(c). The folding pathway is tuned by short single-stranded DNA molecules that are attached to the long DNA backbone at appropriate positions. In addition, DNA origami can be used in combination with synthetic molecules (colloids and polymers), where the synthetic components can be functionalized. For example, the three legs of the Y-shaped DNA origami in Fig. 14(a,b) can be functionalized with an apolar polymer, so that these macromolecules will be surface active (see the lower panels in Fig. 14(a,b)). Such DNA origamis can be used for surface patterning purposes. Further applications include the

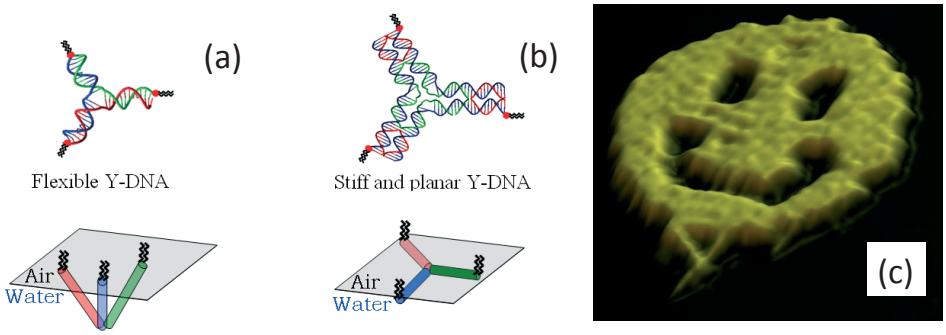


Fig. 14: *(a,b) A Y-shaped DNA origami with functionalized legs, and with (a) a flexible spacer between the three legs, or (b) a stiff, planar structure. The lower panels show the structure of these DNA-hybrids when adsorbed at an air-water interface. Courtesy of Manolis Stiakakis (Forschungszentrum Jülich). (c) A DNA origami scaffold in the form of a "smiley face" (taken from nanotech.org, Paul W.K. Rothemund, Caltech).*

engineering of drug capsules, for example through the construction of artificial virus particles or the construction of DNA tubes, the design of functional materials through self-assembly of DNA-synthetic hybrids, and designing nano-electronic circuitry and plasmonic devices using DNA-origami scaffolds.

4.3 Self-Healing Materials

Self-healing polymers can respond to a traumatic mechanical event (cut, fracture, or strong deformation) to restore the original material and its properties. A variety of architectures have been explored for these materials, from reversible Diels-Alder reactions (formation of covalent $C-C$ bonds) to encapsulation methods. Often these architectures share the problem that the temperature required for the recovery is quite high (around 100 degrees Celsius).

Alternatively there are supramolecular materials, which are naturally self-healing as they are constituted of reversible non-covalent bonds; however, these suffer from low mechanical strength, which drastically limits their applications in structural materials. Materials which contains both supramolecular and covalent bonds are therefore advantageous to tune the properties of the material more towards self-healing or towards mechanical strength, achieving levels for both properties that can be useful for particular applications.

Bio-inspired *hydrogen bonding* is one of the prominent key factors for the design of novel self-healing properties. These bonds counteract damage processes directly and autonomously on the molecular level. The hydrogen-bonding mechanism may be specifically implemented in a network, so that in addition to the usual covalent chemical cross-linking, a transient but reversible link fraction is also present, as shown in Fig. 15. As a result, a catastrophic failure in materials through crack propagation can be prevented by the self-healing of molecular cracks from the very beginning, i.e. on the molecular level. Furthermore, the stress-relieving dissipative source reduces wear and tear and prolongs the lifetime of such materials.

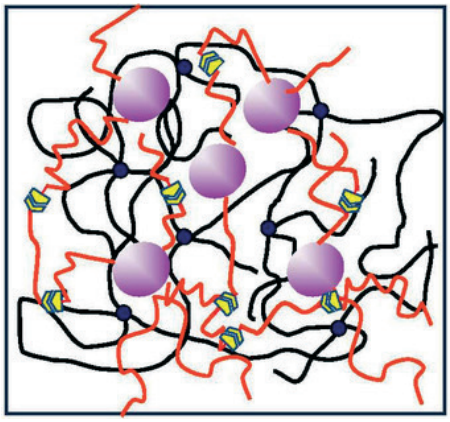


Fig. 15: *Hydrogen-bonds (yellow) serve as transient and reversible bonds in a network, in addition to covalent chemical cross-links (black dots).*



Fig. 16: *(Left) Water droplets do not wet the surface of a Lotus leaf. (Right) This property of super-hydrophobicity is due to a complex structure on the micrometer scale. From Ref. [52].*

4.4 Self-Cleaning Materials

The Lotus leaf has the remarkable property that hardly any dirt, neither oily nor water-soluble, is sticking to its surface, or can easily be removed by rinsing with water, see Fig. 16(a). A closer look at the surface of the Lotus leaf reveals a that it has a hierarchical microstructure [51] with a typical length scale of about 10 μm, see Fig. 16(b). This enables the contact area and the adhesion force between the surface and a water droplet to be significantly reduced, resulting in a self-cleaning process.

When it became clear that self-cleaning qualities come from the physical-chemical properties of super-hydrophobic surfaces at the microscopic to nanoscopic scale, and not from any of the specific chemical properties of the surface of the leaf, this discovery opened up the possibility of using this effect in engineered surfaces [53].

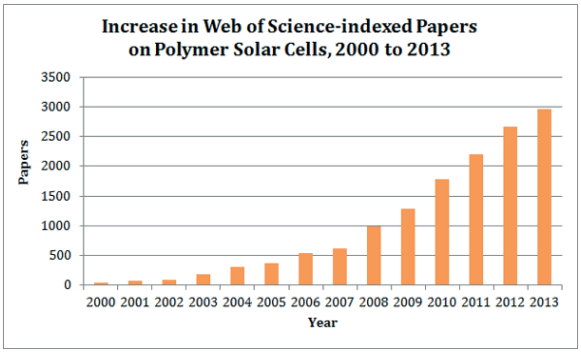


Fig. 17: *Increase of scientific publications on polymer solar cells since 2000. From Ref. [54].*

4.5 Materials for Energy Conversion

Photovoltaic conversion of sunlight is environmentally attractive, and can be done on a domestic scale with silicon cells, the efficiency of which is nudging 20%. However, theres now a lot of interest, see Fig. 17, in a different kind of device to harness sunlight: organic or polymer solar cells (PSCs). These have the potential to provide lightweight flexible films, easy to manufacture, and with the potential to lower costs.

4.6 Biological Materials

Biological macromolecules and biological systems (like cells and organelles) are almost by definition functional – otherwise they would have been singled out by the evolution process. On the macromolecular level, DNA is the storage devise for information, while proteins are the nano-machines which perform the work. Here, a complex network of signaling molecules is responsible for stimulating or depressing the action of chemical reactions. The study of the mechanisms and processes in living systems serves four different goals: (i) to obtain a fundamental understanding of biological processes involving macromolecules on the basis of physical mechanisms and models, and their quantitative analysis by modern physical methods and techniques; (ii) the understanding of the function and dysfunction of proteins and cells and their role in various diseases; (iii) the development of new therapeutic strategies; (iv) the use of the concepts, strategies and mechanisms of biological systems to construct new materials based on bio-macromolecules, bio-synthetic hybrids, or entirely synthetic building blocks.

4.7 Active Matter – Active Materials

In cells, motor proteins generate forces and perform mechanical work. There are several families of such motor proteins, in particular kinesin, dynein, and myosin. Motor proteins perform work by stepping along two types of biological filaments, microtubules and actin. Their functions in the body range from vesicle transport in the cell to muscle contraction, see Fig. 18.
Contrary to passive, diffusive mass transport, active transport can facilitate the movement of a substance also against concentration gradients (from low to high concentrations), possibly towards targeted regions. In cells, such a mass transport towards a region of high concentration is necessary for the accumulating of molecules that the cell needs, such as amino acids, ADP

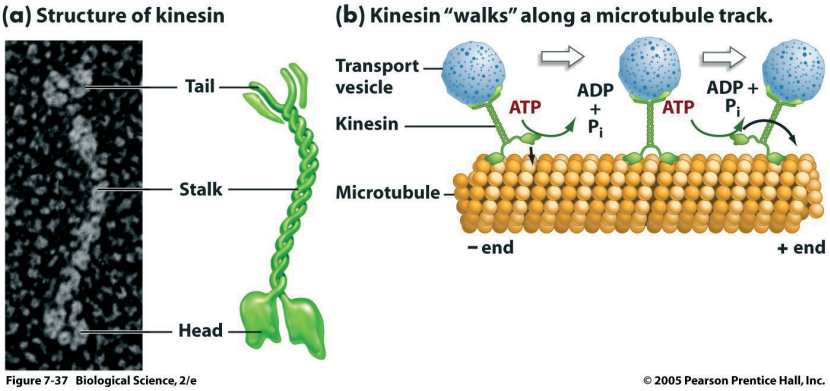


Fig. 18: *(a) The motor protein kinesin consists of two active heads, a passive stalk, and a tail which binds to the cargo). (b) "Walking" occurs by ATP binding to the head, energy release by the reaction* $ATP \rightarrow ADP + P_i$ *and subsequent release of one head from the filament. Transport is directed from the* $-$ *to the* $+$ *end of the polar filament.*

or ATP, glucose, and ions. Examples of active transport are the transport of cargo through molecular motors that move along microtubules, the uptake of glucose in the intestines and of ions in root hair cells of plants.

Motor proteins can also be employed to generate force in artificial systems. One example are *active gels*, which essentially consist of the biological filaments and motor proteins mentioned above, plus ATP as fuel, but nothing else. In this case, motor mini-filaments are used, which consist of a few motor proteins linked together, so that they can bind to two actin filaments or microtubules simultaneously, and displace them relative to each other. This leads to active structure formation, in which bundles, asters, or vortices can form depending on ATP concentrations and filament density. In a dense system of long flexible filaments, motor activity can be used, for example, to speed up the relaxation process from the slow reptation in passive polymer melts to a liquid-like flow with active motors [55]. New features can be obtained by adding permanent cross-linkers, see Fig. 19. In this case, motor activity leads to local contraction and bundle formation in the network, e.g. see Ref. [56].

Another important type of active matter are microswimmers and self-propelled particles. Many uni- and multi-cellular organisms use hair-like filaments (called flagella) driven by motor-protein complexes to propel themselves through a fluid. Locomotion on the nanoscale through a fluid environment is one of the grand challenges confronting nanoscience today [57]. The vision is to synthesize, probe, understand, and utilize a new class of motors made from nanoscale building blocks that derive on-board or off-board power from in-situ chemical reactions. The generated mechanical work allows these motors to move through a fluid phase while simultaneously or sequentially performing a series of tasks. A large variety of such swimmers have been constructed recently, from bimetallic nanorods and Janus colloids to artificial sperm and artificial bacteria.

An example is the screw-like actuated nano- and micro-propellers [58] displayed in Fig. 20, which are rotated by an external magnetic field – almost self-propelled particles, but not quite yet. Another example is artificial cilia, which form spontaneously in a carpet of microtubules grown on a solid substrate, when synthetic polymers are added as depletant to induce an at-

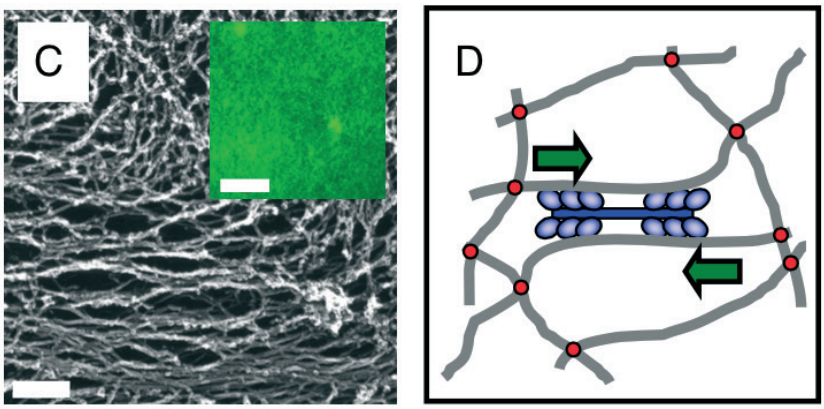


Fig. 19: *(Left) Active networks are macroscopically homogeneous, but small contractile foci are present (inset); the actin filaments extending out from these foci are bundled and aligned, but they merge into a more isotropic network between the foci. (Right) Proposed mechanism of active stiffening: myosin mini-filaments (blue) contract actin filaments (gray) toward one another, thereby pulling against the cross-links (red) and generating an internal stress. From Ref. [56].*

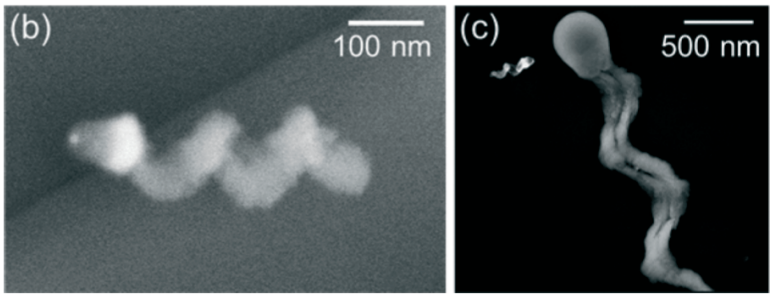


Fig. 20: *(Left) Nanohelix with Au-head and Ni-screw. (Right) Size comparison between nano- and micro-screws. From Ref. [58].*

tractive interaction and thus the formation of microtubule bundles, as well as motor-protein mini-filaments to move the microtubules in a bundle relative to each other [59].

5 Grand Challenges

The importance of research in functional soft-matter systems has been recognized by the *Federation of American Scientists* already in 2003, when the goals of "Understanding of Complex Systems", "Applying Physics to Biology and Medicine", and "New Materials" were identified as three of seven Grand Challenges in Physics for the 21st century [60]. A similar conclusion was reached by a committee of the *National Research Council (USA)* in a 2007 study: "What are the prospects for Condensed-Matter and Materials Physics in the early part of the 21st century?", where "What is the physics of life?" and "What happens far from equilibrium and why?" were identified as two out of six Grand Challenges [61].
Furthermore, the 2012 report From Quanta to the Continuum: Opportunities for Mesoscale Science of the *Basic Energy Sciences Advisory Committee* for the US Department of Energy [62] describes the visionary goals of mesoscale science for future energy-related material development:

- The ability to manufacture at the mesoscale, i.e. the directed assembly of mesoscale structures that yields cheaper, higher performing, and longer lasting products with unique functions.
- The realization of biologically inspired complexity and functionality with inorganic earth-abundant materials to transform energy conversion, transmission, and storage.
- The transformation from top-down design of materials and systems with macroscopic building blocks to bottom-up design with nanoscale functional units producing next-generation technological innovation.

These challenges and opportunities do not only concern energy-related materials, but also many other fields of material sciences to the same extent. Furthermore, it is important to emphasize the aim to realize biologically inspired complexity and functionality in functional materials of the future.

References

[1] W. B. Russel, *The Dynamics of Colloidal Systems* (University of Wisconsin Press, London, 1987).

[2] W. B. Russel, D. A. Saville, and W. R. Schowalter, *Colloidal Dispersions* (Cambridge University Press, Cambridge, 1989).

[3] T. G. M. van de Ven, *Colloidal Hydrodynamics* (Academic Press, London, 1989).

[4] R. J. Hunter, *Foundations of Colloid Science* (Clarendon Press, Oxford, 1989).

[5] J. K. G. Dhont, *An Introduction to Dynamics of Colloids* (Elsevier, Amsterdam, 1996).

[6] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience Publishers, New York, 1969).

[7] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).

[8] J. D. Ferry, *Viscoelastic Properties of Polymers* (Wiley, New York, 1980).

[9] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).

[10] J. de Cloizeaux and G. Jannink, *Les Polymères en Solution: leur Modélisation et leur Structure* (Les Éditions de Physique, Les Ulis, France, 1987).

[11] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (Springer-Verlag, New York, 1997).

[12] I. W. Hamley, *Block Copolymers in Solution: Fundamentals and Applications* (Wiley, Chichester, 2005).

[13] J. Meunier, D. Langevin, and N. Boccara, eds., *Physics of Amphiphilic Layers* (Springer-Verlag, Berlin, 1987).

[14] W. M. Gelbart, A. Ben-Shaul, and D. Roux, eds., *Micelles, Membranes, Microemulsions, and Monolayers* (Springer-Verlag, Berlin, 1994).

[15] G. Gompper and M. Schick, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. Lebowitz (Academic Press, London, 1994), vol. 16, pp. 1–176.

[16] S. A. Safran, *Statistical Thermodynamics of Surfaces, Interfaces, and Membranes* (Addison-Wesley, Reading, MA, 1994).

[17] R. Lipowsky and E. Sackmann, eds., *Structure and dynamics of membranes - from cells to vesicles*, vol. 1 of *Handbook of biological physics* (Elsevier, Amsterdam, 1995).

[18] P. Kumar and K. L. Mittal, *Handbook of Microemulsion Science and Technology* (Marcel Dekker, New York, 1999).

[19] D. Nelson, T. Piran, and S. Weinberg, eds., *Statistical Mechanics of Membranes and Surfaces* (World Scientific, Singapore, 2004), 2nd ed.

[20] H.-G. Elias, *Makromoleküle* (Hüthig & Wepf, Basel, 1990).

[21] J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1992).

[22] D. F. Evans and H. Wennerström, *The Colloidal Domain, where Physics, Chemistry, Biology and Technology Meet* (VCH Publishers, New York, 1994).

[23] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 1995).

[24] M. E. Cates and M. R. Evans, *Soft and Fragile Matter: Nonequilibrium Dynamics, Metastability and Flow* (Institute of Physics, London, 2000).

[25] J. K. G. Dhont, G. Gompper, and D. Richter, eds., *Soft Matter — Complex Materials on Mesoscopic Scales*, vol. 10 of *Matter and Materials* (Forschungszentrum Jülich, Jülich, 2002).

[26] G. Gompper, U. B. Kaupp, J. K. G. Dhont, D. Richter, and R. G. Winkler, eds., *Physics meets Biology — From Soft Matter to Cell Biology*, vol. 19 of *Matter and Materials* (Forschungszentrum Jülich, Jülich, 2004).

[27] T. A. Witten and P. A. Pincus, *Structured Fluids: Polymers, Colloids, Surfactants* (Oxford University Press, Oxford, 2006).

[28] W. C. K. Poon and D. Andelman, *Soft Condensed Matter Physics in Molecular and Cell Biology* (Taylor & Francis, Boca Raton, 2006).

[29] I. W. Hamley, *Introduction to Soft Matter: Synthetic and Biological Self-Assembling Materials* (Wiley, Chichester, 2007).

[30] J. K. G. Dhont, G. Gompper, G. Nägele, D. Richter, and R. G. Winkler, eds., *Soft Matter — From Synthetic to Biological Materials*, vol. 1 of *Key Technologies* (Forschungszentrum Jülich, Jülich, 2008).

[31] J. K. G. Dhont, G. Gompper, P. Lang, D. Richter, M. Ripoll, D. Willbold, and R. Zorn, eds., *Macromolecular Systems in Soft and Living Matter*, vol. 20 of *Key Technologies* (Forschungszentrum Jülich, Jülich, 2011).

[32] G. Gompper and M. Schick, eds., *Soft Matter – Polymer Melts and Mixtures*, vol. 1 (Wiley-VCH, Weinheim, 2005).

[33] G. Gompper and M. Schick, eds., *Soft Matter – Complex Colloidal Dispersions*, vol. 2 (Wiley-VCH, Weinheim, 2005).

[34] G. Gompper and M. Schick, eds., *Soft Matter – Colloidal Order: Entropic and Surface Forces*, vol. 3 (Wiley-VCH, Weinheim, 2007).

[35] G. Gompper and M. Schick, eds., *Soft Matter – Lipid Bilayers and Red Blood Cells*, vol. 4 (Wiley-VCH, Weinheim, 2008).

[36] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, Sunderland, MA, 2001).

[37] D. H. Boal, *Mechanics of the Cell* (Cambridge University Press, Cambridge, 2002).

[38] H. Schiessel, J. Phys. Condens. Matter **15**, R699 (2003).

[39] R. Phillips, J. Kondev, J. Theriot, and N. Orme, eds., *Physical Biology of the Cell* (Garland Science, Taylor & Francis, New York, 2008).

[40] E. Sackmann and R. Merkel, *Lehrbuch der Biophysik* (Wiley-VCH, Weinheim, 2010).

[41] H. Schiessel, *Biophysics for Beginners: A Journey through the Cell Nucleus* (Taylor & Francis, Boca Raton, 2013).

[42] P. Nelson, *Physical Models of Living Systems* (Macmillan Publishers, New York, 2015).

[43] J. H. Schulman and J. B. Montagne, Ann. N. Y. Acad. Sci. **92**, 366 (1961).

[44] H. T. Davis, J. F. Bodet, L. E. Scriven, and W. G. Miller, in Meunier *et al.* [13], pp. 310–327.

[45] Research in Jülich, Nr. 2/2006, Physics meets Biology.

[46] Http://users.soe.ucsc.edu/ pinal/P1.html (University of California, Santa Cruz).

[47] D. Voet and J. G. Voet, *Biochemistry* (Wiley, New York, 1995), 2nd ed.

[48] K. Müller, D. A. Fedosov, and G. Gompper, Sci. Rep. **4**, 4871 (2014).

[49] N. C. Seeman, Nano Lett. **1**, 22 (2001).

[50] P. W. K. Rothemund, Nature **440**, 297 (2006).

[51] W. Barthlott and C. Neinhuis, Planta **202**, 1 (1997).

[52] Http://en.wikipedia.org/wiki/Lotus_effect.

[53] M. L. Ma and R. M. Hill, Curr. Opin. Colloid Interface Sci. **11**, 193 (2006).

[54] Simon Milton, *Polymer Solar Cells: Bringing the Sun's Energy down to Earth*, Science-Watch March 2014. http://sciencewatch.com/articles/polymer-solar-cells-bringing-suns-energy-down-earth.

[55] D. Humphrey, C. Duggan, D. Saha, D. Smith, and J. Käs, Nature **416**, 413 (2002).

[56] G. H. Koenderink, Z. Dogic, F. Nakamura, P. M. Bendix, F. C. MacKintosh, J. H. Hartwig, T. P. Stossel, and D. A. Weitz, Proc. Natl. Acad. Sci. USA **106**, 15192 (2009).

[57] G. A. Ozin, I. Manners, S. Fournier-Bidoz, and A. Arsenault, Adv. Mater. **17**, 3011 (2005).

[58] D. Schamel, A. G. Mark, J. G. Gibbs, C. Miksch, K. I. Morozov, A. M. Leshansky, and P. Fischer, ACS Nano **8**, 8794 (2014).

[59] T. Sanchez, D. Welch, D. Nicastro, and Z. Dogic, Science **333**, 456 (2011).

[60] FAS Public Interest Report, The Journal of the Federation of American Scientists, Volume 56, Number 2 (Summer 2003). http://www.fas.org/faspir/2003/v56n2/challenge.htm.

[61] FYI: The AIP Bulletin of Science Policy News, Number 63: June 21, 2007. http://www.aip.org/fyi/2007/063.html.

[62] *From Quanta to the Continuum: Opportunities for Mesoscale Science*, report for the Basic Energy Sciences Advisory Committee, Mesoscale Science Subcommittee. http://science.energy.gov/ /media/bes/pdf/reports/files/OFMS_rpt.pdf.

MATERIALS



A 1 Functional Nanoparticles

A. M. Schmidt
Chemistry Department
Universität zu Köln

Contents

1 Introduction 2

2 Size-dependent Physical Phenomena in Nanoparticles 2
2.1. Surface Area-to-Volume Ratio 2
2.2. Localized Surface Plasmon Resonance 4
2.3. Semiconductivity and Fluorescence 6
2.4. Magnetism & Related Phenomena 7

3 Synthesis of Fuctional Nanoparticles 10
3.1 Anisotropic & Multicomponent Nanoparticles 11
3.2 Stabilization, Compatibilization 12

4 Examples for Functional Nanoparticle Systems 13
4.1 Nanoparticles as Mechanical Strengtheners & Stimulators 13
4.2 Optical Detection & Diagnostics 15
4.3 Local heating 16
4.4 Multifunctional 17

References 18

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

Inorganic nanocrystalline materials exhibit exeptional structural, mechanical and other physical properties that are of high interest for application in today´s key technogies in biomedics, optical electronics, and data storage [1]. The developmemt of new methods to understand and analyze the underlying nanoscale and interface effects is of critical importance for the future development of this emerging field of materials science and application. The advent of nanoscience and –technology offers the possibility of creating novel model nanostructures with precisely controlled composition, structure and morphology, and thus with a previously inaccessible control of properties [2].

This lecture gives an overview on the variety of fuctional inorganic nanoparticles and their size-dependent optical, dielectric and magnetic properties that are useful for the investigation of the underlying fundamental properties and for implementation into nanostructured multifunctional materials and systems.

2 Size-dependent Physical Phenomena in Nanoparticles

It is well known that the decrease of crystal size and the large surface area present in inorganic nanoparticles result in a spectrum of size-dependent physical properties that are of great interest to explore, to understand, and to tailorize for future high-performance and multifunctional materials. Modern nanotechnology offers a number of functional metallic and other inorganic nanomaterials with a property spectrum that is highly dependent of the particle size and shape, as well as of the surface functionalization and chemical environment. As a consequence of their optical, dielectrical and magnetic characteristics originating from collective electron responses in crystalline structures (in combination with limited crystal size and surface dominance), such functional nanoparticles are of high interest for sensing, catalytic, or energy-transducing applications.

2.1. Surface Area-to-Volume Ratio

In general, when an object gets smaller, its surface area-to-volume ratio (A/V) ratio increases. For a compact geometry like a cube or sphere with a characteristic length r, $A/V \sim 1/r$ (see Fig. 1:. An increase in A/V implies the exposure of a larger fraction of atoms to the environment. For a 10 nm structure, about 30 % of all builing blocks are located at the surface of the nanocrystal, and for nanoporous, tubular and ribbon-like structures, this ratio is even higher.

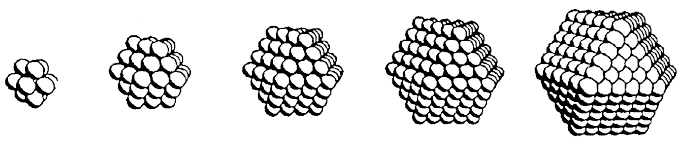
It is obvious that the number of nearest neighbours within the crystal lattice is reduced for atoms that are located near the crystal surface as compared to bulk atoms. This is of consequence not only for the direct chemical environment of the atom, but also correspondingly induces a redistribution of electronic charge, altering the binding situation [3]. Similarly, atoms located near the interface of two materials experience a different local chemical and electronic environment than atoms in the bulk of the materials. As a result, the energy of these atoms is different from that of the atoms in the bulk. What is more, for various metal-based nanoparticles, larger values of surface-excess free energy are found as compared to the bulk material, originating from a size-dependent lattice parameter (SDLP) as found in many metals [4-7].

For particles in the nanometer range, this leads to an excess surface free energy comparable to the overall lattice energy, resulting in a general structural instability, and predisposition for agglomeration and Ostwald-type ripening. It is therefore generally necessary to stabilize

nanoparticles against these aging processes to result in well-defined systems with reliable properties. At the same time, the increased activity of surface-bound atoms as compared to neighbors located in the bulk is beneficial for the implementation of specific surface binding sites, e. g. for coupling to the environment, catalysis, and adsorption.
First of all, the large ratio of surface area to volume in nanosized objects makes the behavior of surfaces and interfaces become a prominent factor controlling the mechanical properties of nanostructured materials [8-10]. In the case of polymeric nanocomposites, the elastic properties of the interface region between nanoscopic fillers and the polymer matrix are predominant for the stress–strain relationship and need to be given due consideration while discussing their overall properties. The effective interfacial properties strongly depend on the texture of the particle surface, its chemical composition, and the presence of surfactants and or charge carriers. Imperfections and gradual composition profiles gives rise to the concept of an "interphase" that corresponds to the volume defined by the narrow region sandwiched between the two phases but with different properties.

In the absence of strong covalent or coordinative interactions between the particle surface and the matrix components (e. g., polymers), such nanocomposites can be considered to be a Class I hybrid material. In this case, the interaction between the inorganic filler particles and the matrix is dominated by physical forces, including electrostatic interactions, van-der-Waals forces, and surface complex or hydrogen bond formation [11,12]. In contrast, Class II nanocomposites involve even stronger, generally covalent interactions between particle surface and matrix phase. Recently, several methods have been reported on the modification of inorganic components with polymerizable groups, that subsequently can be copolymerized with organic monomers [13-16].

The high surface area and easy modification of the particulate surface properties present in nanoparticles is promising for applications in quasi-homogeneous catalytic systems. The nancrystalline catalysts can be suspended in organic or aqueous media as required after appropriate compatibilization. In numerous cases, and enhanced surface-normalized activity and modified selectivity has been observed in nanosized catalysts as compared to larger particles. Similarly, small particles are generally attractive for tailored adsorption processes with high accessible surface area.



Full-shell "magic number" clusters					
Number of shells	1	2	3	4	5
Number of atoms in cluster	13	55	147	309	561
Percentage of surface atoms	92	76	63	52	45

Fig. 1: *Idealized representation of full-shell metal clusters with "magic numbers" of atoms, based on the dense sphere packing. Reprinted with permission from Elsevier*, *J. Mol. Catal. A* 1999 [76].

2.2. Localized Surface Plasmon Resonance

The general optical properties of metal nanoparticles differ significantly from those of the respective bulk materials. In the sub-100 nm regime, the continuous electron bands become discrete, and properties such as plasmon resonance scale with particle size and are strongly influenced by the particle environment and state of aggregation [17,18].
Surface plasmon resonance originates from the metallic electrons in the conducting materials delocalized within the crystal region. In the bulk transition metals, the mean free path of the d-type conduction electrons is in the range of a few tens of nm. For objects with a characteristic length scale below this value, no scattering is expected from the bulk phase of the particle, and all interactions with electromagnetic waves are expected to be with the surface.

Localized surface plasmon resonance (LSPR) is an optical phenomenon generated by light when it interacts with conductive nanoparticles that are smaller than the incident wavelength. As in surface plasmon resonance, the electric field of the incident light collectively excites electrons of a conduction band, inducing a displacement of the electron cloud relative to the nuclei and giving rise to a restoring force from coloumbic attraction. The resulting oscillation of the electron cloud with respect to the nuclei meets the resonance condition at a specific frequency with electrons exhibiting a $\pi/2$ phase lag with respect to the field of incident radiation. The resonantly enhanced field inside the nanoparticle leads to a dipolar field on the exterior that is responsible for enhanced absorption and scattering cross-sections, as well as strongly enhanced EM fields in close vicinity of the nanoparticle surface [19]. The resonance frequency depends on the composition, size, geometry, the dielectric environment and the separation distance of the nanoparticles. Resonance of the incident light with the oscillation results in a strong absorption, causing the intense colour of metallic nanoparticles [20].

Although this phenomenon has always been fascinating to scientists, it was only during the recent twenty years that synthetic methods became available allowing to accurately manipulate the structure of inorganic objects on the nano- and mesoscale. In consequence, many applications became possible that take advantage of the strong electric field enhancement on the particle surface.

When only dipole oscillations contribute to the extinction cross-section C_{ext}, Mie's solution of the Maxwell equation can be used to obtain the spectrum for well-separated nanoparticles [21,22]:

$$C_{ext} = \frac{24\pi^2 R^3 \varepsilon_m^{3/2} N}{\lambda \ln(10)} \frac{\varepsilon''}{\left(\varepsilon' + \chi\varepsilon_m\right)^2 + \left(\varepsilon''\right)^2} \qquad (1)$$

with ε_m: dielectric constant of the surrounding medium, $\varepsilon = \varepsilon' + i\varepsilon''$: complex dielectric constant of the bulk metal, R: radius of the nanoparticle, N: electron density.
The factor χ accounts for the shape of the particle. It is assigned a value of two for a spherical particle, and can be as large as 20 for particles with high aspect ratios such as nanorods [23]. The surface plasmon resonance peak responsible for the color of a spherical NP is observed when $\varepsilon' = -2\ \varepsilon_m$. Using extensions of the model, the extinction cross-section for NPs of various shapes can be theoretically modeled [24].

While the absorption coefficient caused by the plasmon resonance of spherical metal nanoparticles is already orders of magnitude larger than for strongly absorbing dyes [25], the effect can even be enhanced for particles with elongated or anisotropic shapes. As the shape or size of the nanoparticle changes, the associated change in surface geometry causes a change in the oscillation frequency of the electrons, generating different cross-sections for the optical properties including absorption and scattering. The anisotropy has been shown to generate large control over the optical absorbance for all shapes generated [26]. Nanorods possess two plasmon resonance bands [27], one due to the transverse oscillation of the electrons (around 520 nm for gold), and the other due to the longitudinal plasmon resonance at longer wavelengths as shown for various aspect ratios in Fig. 2a. Disks also display a similar plasmon resonance absorption dependence on their aspect ratio [28], while in triangular nanoparticles the edges and corners play an important role [29].

Similarly, a change in the dielectric constant of the surrounding material is of effect on the resonance frequency, as it alters the electron charge density accommodation at the interface. Accordingly, metal nanoparticles can serve as highly sensitive transducers of small changes in the local refractive index, indicated by a spectral shift of extinction (absorption plus elastic light-scattering) and scattering spectra. The colour shift can be achieved by a change of the dispersion medium, but is particularly sensitive to the immediate neighborhood to the crystal surface, i. e. the capping agent or stabilizer shell and its thickness, composition and electronic properties.

For many organic molecules with a relatively high refractive index compared to solvent or air, their immobilization on the nanoparticle surface results in a redshift of the particle spectrum. There is an empirical releaionship between the magnitude of the spectral shift of LSPR extinction (or the scattering wavelength maximum, respectively) $\Delta\lambda_{max}$, for small nanoparticles [30]:

$$\Delta\lambda_{\max} = m\Delta n\left[1 - e^{\frac{-2d}{l_d}}\right] \tag{2}$$

with m: bulk refractive index response of the nanoparticle (sensitivity factor), Δn: change in refractive index, d: effective thickness of the adsorbed layer, l_d: characteristic field decay length. This expression serves as the basis for using the LSPR wavelength-shift in chemical assays and bioassays to probe the (bio)chemical affinity between molecules, or as highly sensitive optical sensors in diagnostics and trace analysis. Examples for such applications are found in chapter 3 [18].

Common nanomaterials used for LSPR applications are noble transition metals such as Ag and Au. The d – d transition energy levels lead to LSPR in the visible range of the spectrum [31], and the electrochemical stability is of advantage for particle handling and reliable properties. Although Ag exhibits the sharpest and strongest bands among all metals, Au is preferred for biological applications due to its inert nature and biocompatibility [32], and thiol-gold association for immobilization of biomolecules.

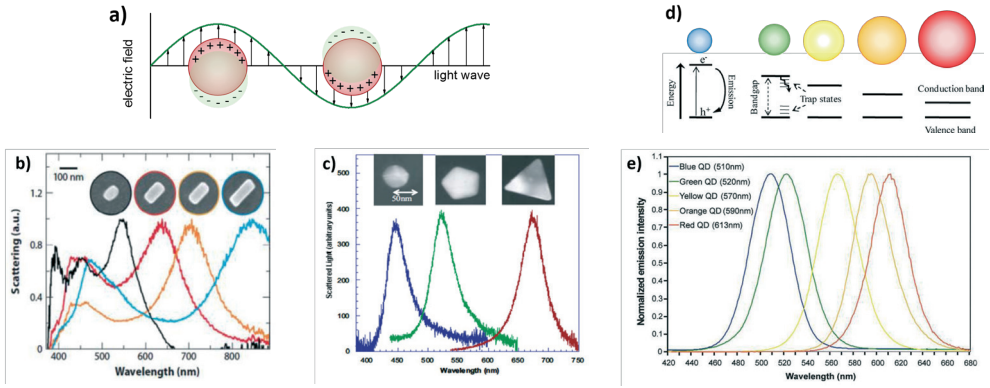


Fig. 2: *a) Illustration of the LSPR principle in metal nanoparticles, and shape-dependent scattering spectra of (b) rod-like silver nanoparticles with different aspect ratio (reprinted with permission from Annual Reviews Publishers, Ann. Rev. Phys. Chem., 2009* [36]*), and (c) of single silver nanoparticles of different shapes obtained in dark-field configuration (reprinted with permission from AIP Publishing, J. Chem. Phys. 2002* [37]*). The insets are SEM images of the respective particles. d) Illustration of the size-dependent electronic states in quantum dots* [35]*, and d) emission spectra of water-soluble CdSe/ZnS QDs excited at 350 nm (reprinted with permission from Macmillan Publishers Ltd., Nature Biotechnology 2002* [34]).

Unlike organic fluorophores, plasmonic NPs do not photobleach or blink and thus can serve as intense and robust labels for biosensors, immunoassays, cellular imaging and surface-enhanced spectroscopies [33]. For larger particles, modes with higher multipoles can occur, and in practical embodiments, the particle polarizability has to be considered.

2.3. Semiconductivity and Fluorescence

Semiconductor particles with sizes between about 1 nm and 10 nm have received much attention since the pioneering works in the early 1980s [38]. They are well-known for their strong band-gap luminescence tunable by size as a result of the quantum confinement effect, being of interest for different applications, such as fluorescence labels, as optical amplifier, or in thin film electroluminescent devices [42].

Semiconducting properties of macrocrystalline semiconductors based on Cd chalcogenites, ZnO, and others, originates from the spatial overlapping of orbitals from periodically arranged atoms in a crystalline lattice, resulting in a spanning valence- and conduction band of orbitals. Due to attractive Coulomb forces, an exited electron and the corresponsing hole can form a Wannier-type exciton delocalized within the bulk crystal lattice. In a nanoparticle, the particle size becomes comparable to or is even smaller than the exciton diameter in the bulk material. As a consequence, the continuous energy bands split into discrete levels and the electronic excitation shifts to higher energy values. Using the so-called effective mass approximation, the size dependence of the energy of the first excitonic transition (or the band gap shift with respect to the bulk value E_g) on the particle radius R can be calculated as [46]:

$$E(R) - E_g = \frac{h^2}{16R^2}\left[\frac{1}{m_e^*} + \frac{1}{m_h^*}\right] - \frac{1.8e^2}{\varepsilon R} \quad (3)$$

with $E(R)$ particle radius R depending band gap, m_e^* and m_h^* effective mass of hole and electron, respectively, ε: dielectric constant, where the first term in Eq. (3) represents the particle-in-a-box quantum localization energy ($1/R^2$ dependence) and the second term the Coulomb energy ($1/R$ dependence). The major consequence of the quantum confinement effect is an increase of the apparent band gap with decrease of the particle size R, experimentally observed as a blue shift as well in the absorption onset as for the emission spectra (Fig. 2b). The effect is very sensitive to small changes in crystal size, and a broad spectrum of emmission colours can be achieved in nanocrystals below 10 nm with high quantum yields [43]. Within this size range, an effective tailoring of the optical properties of semiconductors is possible simply through control of their size and the size deviation, which can be done either at the stage of synthesis or in postpreparative stages. Depending on the chemical and electronic structure of the surface absorbants, the structure and morphology of an absorption layer may have a drastic effect of the otical and electronic structures observed in metallic nanostructures. The resulting band bending observed in macro- and mesoscopic semiconductor layers and wires can result in a significant influence on the cross-section for electron conduction and optics.
The drastic changes in optical properties governed by the particle size and the immediate particle environment inspired scientist to develop advanced synthetic routes to nanoparticles of different semiconductor materials cadmium and other metal chalcogenides, and protocols for their use in advanced sensor and optical applications.

2.4. Magnetism & Related Phenomena

The behavior of nanoparticles based on ferromagnetic metals such as cobalt, iron, or nickel, and the ferrimagnetic iron oxides and ferrites is dominated by their response to external magnetic fields, including pattern formation and magnetic heating.

Macroscopic crystalline magnetic materials tend to split up into magnetic domains, each characterized by the parallel alignment of the involved magnetic moments on the atomic scale within the domain. In the absence of external fields, the relative arrangement of the domains is organized to minimize the net overall magnetization. For nanocrystals, however, the formation of magnetic domains is energetically not favored, and the particles can be described as single-domain particles (SDP). The critical diameter for domain formation can be approximated at 150 nm for Fe_3O_4 and at 70 nm and 14 nm for Co and Fe, respectively [47]. When a single-domain magnetic particle is well-dispersed in a suitable carrier fluid, it can be rationalized as a magnetic dipole that can reorientate by two mechanisms: either by rotation of the whole particle within the carrier (Brownian type) or by collective reorientation of the atomic magnetic moments within the crystal lattice of the particles (Néel type, see below). In either way, the thermal fluctuation of the dipoles leads to a zero net magnetization in the absence of outer fields. Under field and gradient influence, an alignment of the dipoles is achieved, while a suitable stabilization prevents the particles from (irreversible) agglomeration. As a consequence, magnetic colloids can be moved in droplets or set into position against gravity by the influence of outer fields.

In homogeneous static fields, the degree of alignment of an ensemble of monodisperse, non-interacting magnetic nanoparticles depends on the field strength and follows the Langevin law [48] for paramagnetic substances:

$$M = M_s\left[\mathrm{ctgh}(\alpha) - 1/\alpha\right] \tag{4}$$

with M: sample magnetization, M_s: saturation magnetization of the sample, and

$$\alpha = \frac{\mu_0 mH}{k_B T} \tag{5}$$

with μ_0: permeability of free space; m: magnetic moment of the particles; H: magnetic field strength; k_B: Boltzman constant; T: temperature. While the general behaviour of magnetic nanoparticle dispersions thus resembles that of paramagnetic materials, the involved magnetic moments of the particles are in the order of $10^3 - 10^4$ μ_B. This behavior is therefore referred to as superparamagnetic, and the critical size for a stable magnetization at room temperature is known as the superparamagnetic limit. As a consequence, the magnetization curve of magnetic colloids shows a symmetrical sigmoidal shape with no hysteresis [48]. More complex behaviour is observed when the particles are interacting, immobilized in a solid or viscous matrix, or organized in a 2D or 3D array.

In dynamic magnetic fields, the different relaxation times of the Néel type remagnetization mechanism and the Brownian type becomes important [49]. Both processes are thermally activated. In case of the "inner" (Néel type) remagnetization, this requires activation against the magnetic anisotropy barrier $K_{\mathrm{eff}}V$ (with K_{eff}: effective magnetic anisotropy constant of the particle, and V: particle volume). The magnetic anisotropy is mainly composed of the magnetocrystalline anisotropy of the magnetic component, and the shape anisotropy. In some cases, significant contribution from surface effects and magnetoelastic effects have to be taken into account.

$$\Delta E_N = K_{eff} V_c = K_{eff} \frac{4}{3}\pi R^3 \tag{6}$$

with a Néel relaxation time τ_N of

$$\tau_N = \tau_0 e^{\frac{\Delta E_N}{k_B T}} \tag{7}$$

and $f_0 = 1/\tau_0$: Lamour frequency of the magnetization vector, typically in the range of 10^{-9} s^{-1} for superparamagnetic nanoparticles. Similarly, the Brownian relaxation is counteracted by the friction acting on the particle, that is, approximately against the viscosity η_0 of the carrier fluid:

$$\tau_B = \frac{4\pi\eta_0 R^3}{k_B T} \tag{8}$$

with τ_B: Brownian relaxation time. While τ_N depends exponentially on the core volume, τ_B increases linearly with the hydrodynamic volume of the particles. The latter may differ from

the volume of the crystalline cores in the case of a stabilizing shell or agglomeration. The faster process dominates the overall remagnetization process, and for similar time scales, both mechanisms contribute according to $1/\tau_{eff} = 1/\tau_N + 1/\tau_B$.

The dynamic magnetic losses can be detected by AC susceptomery. Here, the time-depending response of the sample magnetization is recorded in a frequency-modulated magnetic field of low amplitude. The real and and imaginary part of the initial susceptibility, χ' and χ'' is determined, and a peak in the phase shift $\tan\delta = \chi''/\chi'$ is observed in the region of τ_{eff}. In real particle ensembles, an influence of the particle size distribution has to be considered.

$$\chi(\omega) = \chi_o \int \frac{1}{(1 + i\omega\tau_B(R_h))} f(R_h) dR_h \tag{9}$$

In recent years, the comlex properties of magnetic nanostructures have been intensively investigated, motivated by potential technological and medical applications as recording media, magnetic sensors and biocompatible tools for diagnostics and therapy (see chapter 3).

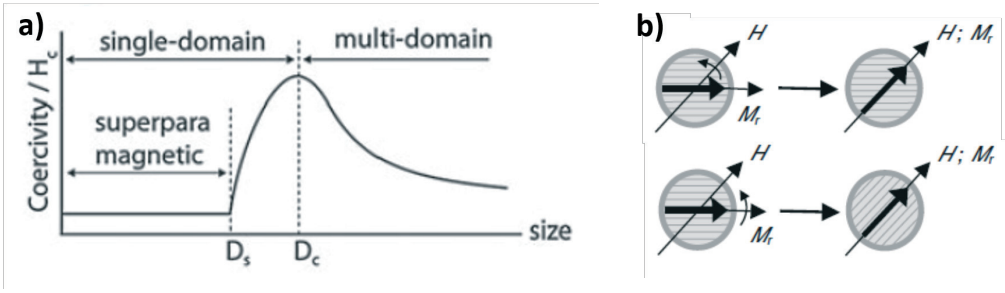


Fig. 3: *a) Illustration of the coercivity dependence on the particle diameter for magnetic nanoparticles, b) of the intrinsic (Néel) and extrinsic (Brown) remagnetization mechanism. Reprinted with permission from Springer, Coll. Polym. Sci 2007* [111].

The relaxational and hysteretic losses observed in magnetic nanoparticles when exposed to a dynamic magnetic field of the kilo- to Gigahertz regime lead to a significant generation of heat [53]. Within the validity range of linear response theory, the loss power density P is given by [55]:

$$P(f, H) = \frac{(mHf\tau_{eff})^2}{(2\pi k_B V)(1 + f^2\tau_{eff}{}^2)} \tag{10}$$

with f, H: frequency and field amplitude of the AC field. The magnetic heatability of a magnetic particle ensemble is greatly influenced by the particle size and the size distribution as well as field parameters. The fact that most other material (exept of metals) are transparent in the relevant frequency range makes this option pertinent for local heating applications. Clinical studies are performed regarding the application of the magnetic heatability of magnetite nanoparticles in tumor therapy by magnetic fluid hyperthermia [56].

The full potential of magnetic particles as functional building blocks in nanostructured materials becomes apparent when considering their ability for structure for mation and self

assemby or self organisation with and without the presence of an external field, and their contribution to the formation of novel hybrid materials with anistropic magnetic and mechanical properties (see chapter 3).

3 Synthesis of Fuctional Nanoparticles

The availability of well-defined particle samples with controlled composition, size, shape, and surface characteristics is of crucial importance for the investigation of the structure-property relationships and future applications, requiring high-yield synthesis of nanoparticle ensembles with controlled optical, electronic and/or magnetic properties. A variety of chemical synthesis methods have been developed in the last decade for that purpose. The formation of inorganic nanoscaled particles composed of metals and metal oxides or chalcogenides can principally occur by a bottom–up or a top–down approach [59].
Nanostructured metal colloids accessible by disintegration of macroscopic (bulk) crystals, and subsequent stabilization of the resulting nanosized inorganic particles by the addition of protecting agents [60], e. g. by laser ablation or high performance milling. The properties of the resulting particles, however, are hard to tune, and the control of particle size and dispersibility is anot trivial for these approaches.

For the preparation of magnetic colloids, the bottom–up approach is by far more spread. It is based on the controlled condensation of low molecular precursors such as metal salts or (organo-) metallic complexes either in the gas phase or in solution. Best control of the particle properties is maintained by a controlled nucleation and growth process of the evolving nanoparticles. Alternatively, the growth of the primal particles can be controlled by a confined outer geometry, e.g., in micro- or miniemulsions or in a gel-, respectively glass matrix. Vapor techniques like atom vapor deposition (AVD) and chemical vapor deposition (CVD) have provided chemists with a versatile route for the production of a wide range of nanostructured colloids on a preparative laboratory scale [62]. The use of metal vapor techniques is, however, limited because the operation of the apparatus is demanding and it is difficult to obtain narrow particle size distributions. An alternative are wet-chemical routes by chemical reduction of metal salts, electrochemical path-ways, or the controlled decomposition of metastable organometallic compounds in the presence of stabilizing agent in order to prevent the particle agglomeration (see chapter 2.2) [59].

Nanoparticle ensembles are generally described as ‘‘monodisperse’’ when possessing a standard deviation s from the mean particle size of than 10 % from the average value, and as havin a narow size distrubution when the de viation of the mean size is about 20%. While principally techniques are available that control the particle size post synthesis by size-selective separation [67], the processes are generally time consuming and difficult to scale up. Size-selective synthetic pathways are therefore preferred.

The pioneering work of Faraday on the chemical reduction of transition metal salts in the presence of stabilizing agents to generate zerovalent metal colloids in aqueous or organic media dates back to 1857 [69], now being the basis for the most common and powerful synthetic methods in this field. Reproducible standard protocols for the preparation of metal colloids were established by Turkevich [70], proposing a mechanism for the stepwise formation of nanoclusters based on nucleation, growth, and agglomeration, that principally is still valid in a refined version. According to this, nanoparticle formation is initiated by the

reduction of the metal salt to give zerovalent metal atoms. These atoms can reversibly collide in solution with other metal ions, atoms, or clusters to form an irreversible nucleus or seed. The diameter of the initial seeds nuclei can be well below 1 nm, depending on the strength of the metal-metal bonds and the difference in redox potentials of the metal salt and the reducing agent applied, and the strength of the metal-metal bond. The seed further grow into bigger particles by coalescence with freshly reduced metal atoms or clusters.

The size and size distrubution of the resulting nanoparticles is controlled by the relative rates of nucleation and particle growth. The essential factors that control the particle size are the strength of the metal – metal bond; the molar ratio of metal salt; colloidal stabilizer; and reducing agent; the extent of conversion or the reaction time; the applied temperature; and the pressure. In practice, the parameters have to be carefully adjusted to the kinetic requirements in order to result in a reproducible, well-controlled synthetic protocol for particles with predermined composition, phase, and size.

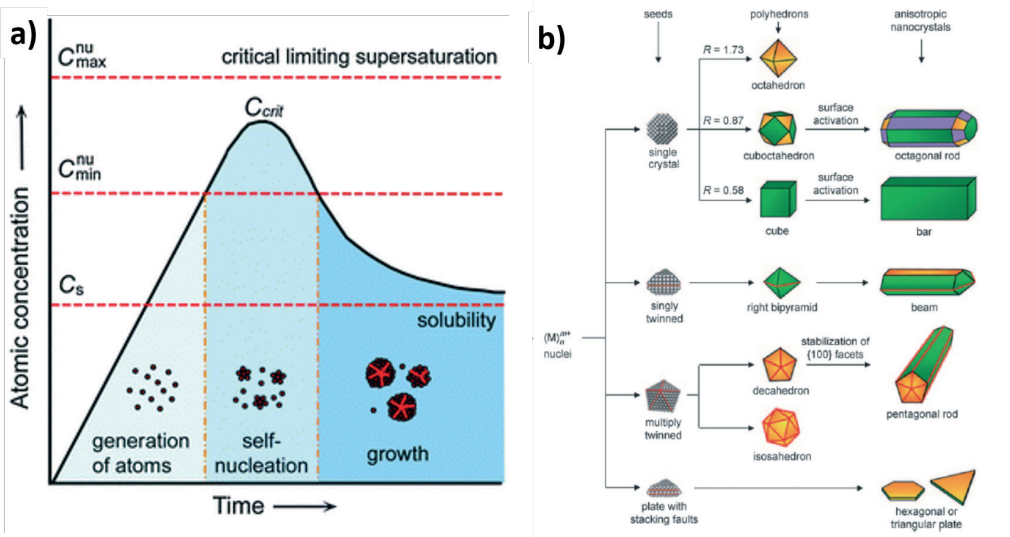


Fig. 4: *a) LaMer model describing nucleation and growth of nanocrystals as a function of reaction time and concentration of precursor atoms. Adapted with permission from ACS, J. Am. Chem. Soc. 1950* [73]*; b) Reaction pathways to* fcc *metal nanocrystals with different shapes; ((100) green; (111) orange, (110) purple; R=<100>/<111>). Reprinted with permission from Wiley, Angew. Chem. Intern. Ed. 2009* [74].

3.1 Anisotropic & Multicomponent Nanoparticles

The properties of a metal nanocrystal are determined by a set of physical parameters that may include its size, shape, composition, and structure (e.g., solid or hollow). In principle, one can tailor and fine-tune the properties of a metal nanocrystal by controlling any one of these parameters, but the flexibility andbscope of change are highly sensitive to the specific parameter.
For example, in the case of localized surface plasmon resonance (LSPR) and surface-enhanced Raman scattering (SERS), both computational and experimental studies have demonstrated that the shape and structure of a Au or Ag nanocrystal play the most important roles in determining the number, position, and intensity of LSPR modes, as well as the spectral region or polarization dependence for effective molecular detection by SERS.

An important research direction in the synthesis of functional nanoparticles is the expansion from simple spherical single-component nanoparticles to anisotropic and/or multi-component hybrid nanostructures with discrete domains of different materials arranged in a controlled fashion [75]. The last decade has witnessed the successful synthesis of metal nanocrystals in a variety of shapes, including spheres, spheroids, cubes; cuboctahedrons; octahedrons; tetrahedrons; right bipyramids; decahedrons; icosahedrons; thin plates with a triangular, hexagonal, or circular profile; and rods or wires with a circular, square, rectangular, pentagonal, or octagonal cross-section [74].

Many methods have been demonstrated for controlling the shape of inorganic nanocrystals, and for wet-chemical methods, the selection of reductant, reaction conditions, and stabilizer are critical to forming a particular shape. Strong reducing agents, as typified by polyol reduction, facilitate the synthesis of thermodynamically favored polyhedral nanocrystals [76].
In multicomponent functional nanoparticles, the different functionalities of two or more components can be combined and integrated, with the dimension and material parameters of the individual components independently optimized [75]. One such example is the Co/CdSe bifunctional magneto-optic nanocrystals reported by Klimov's group [78].
The core/shell nanoparticles retain the magnetic and optical properties of each single component and permit potential applications as optical reporters and magnetic handles for bioassay. Since the dimensions of the individual components are comparable to the size of the biomolecules, the combination is expected to provide improved performance. The second advantage is in providing novel functions not available in single-component materials or structures.

Because many physical and chemical properties have critical length scales on the order of nanometers, the intimate contact between the nanocomponents in hybrid nanoparticles should allow strong interactions between these components and a possible rational modulation of the physical and chemical properties from each individual component. Such composite particles consist of symmetric core/shell nanoparticles [78], non-symmetric heterodimers [83], and other multicomponent heterostructures [87]. There are several comprehensive review articles on the synthesis, properties, and applications of such multicomponent nanostructures [90].

3.2 Stabilization, Compatibilization

Nanoparticulate systems are subjected to a number of destabilizing forces that can lead to phase separation due to agglomeration and sedimentation. One of the main causes of phase separation is the van der Waals interaction between the high-surface particles that results in the formation of agglomerated material [93]. While the sedimentation of single particles by gravitational force is prevented by thermal motion in the size range below 100 nm, the formation of small agglomerates leads to a reduction of the particle diffusion, and sedimentation is accelerated. In addition, the unique size-dependent physical properties of the particles can be affected by agglomerate formation. Consequently, nanoparticle dispersions need to be stabilized against agglomeration by creating a protecting shell for each particle that prevents too close contact, using classical colloid chemistry concepts, including electrostatic stabilization and steric stabilization [65].
Electrostatic stabilization is based on the Coulombic repulsion between equally charged particles. Surface charge is generated by the evolment of an electrical double layer formed by ions adsorbed at the particle surface (or generated by redox reaction at the interface), and the corresponding counterions. This is exemplified by the gold sols prepared by the reduction of

$[AuCl_4]_2$ with sodium citrate, leading to a negatively charged citrate molecules on the particle surface, and a diffuse layer gradient of sodium and citrate atoms around the particle [70].
Steric stabilization is achieved by absorption or coordination of sterically demanding organic molecules such as surfactants or polymers on the particle surface, acting as a protective shell. Coordination or chemisorption is best achieved under apployment of tailored anchors (donor groups), such as phosphanes, amines, thiols or carboxylates; coordinating solvents such as THF or propylene carbonate; long-chain alcohols or surfactants. In general, lipophilic protective agents give particles soluble in organic solvents, while hydrophilic and charged agents are best suited for water-based dispersions. When polymers are involved, they can be either physisorbed or covalently attached to the particle surface, and may be crosslinked to encapsulate the inorganic core. Polymers have the the additional potential to be functionalized to introduce additional functions such as catalytic activity, colour, or biological function.

Apart from compatibility with the environment, the surface properties of the nanoparticles can be of impact on the physical properties of the particles. By interaction between the nanoparticle surface and the matrix or coordination agents (ligands), a pertubation of the electronic and magnetic properties of the crystal lattice close to the interface can be strong enough to result in chemically and/or electronically induced changes in the intrinsic magnetic moments of interfacial atoms and the surface magnetic anisotropy of nanoparticles. The effect is most pronounced for particles an clusters below 10 nm.

4 Examples for Functional Nanoparticle Systems

Owed to their unique physical properties, functional inorganic nanoparticles are used in a broad range of applications, including responsive materials for sensing and actuation and biomedical applications. Their combination with other inorganic or organic compounds to hybrid particles or materials allows the composition of structures with a property spectrum that is otherwise not accessible.

4.1 Nanoparticles as Mechanical Strengtheners & Stimulators

A classical application of inorganic nanoparticles is their use as mechanical fillers in polymeric matrices. In general, the fillers improve mechanical properties of polymer materials, e.g. they enhance the different mechanical moduli, and ultimate strength of the material. The smaller the size of the particles, the larger is the interface where interactions between polymer molecules and fillers can generate new properties.
In filled polymers, the mobility of the macromolecules is often reduced as compared to the non-filled material. This is ascribed to a confinement effect by the filler particles, or due to interactions between filler as matrix molecules. This effect is usually more pronounced for nanometer sized fillers than for micrometer sized fillers. For several systems, the formation of of layered interphases of slowed molecular dynamics around the particles is reported [97]. In elastomers, interphases with almost immobile molecules are known as bound rubber. For a decent review see [100].
Among the most frequently employed inorganic filler particles is SiO_2 and carbon black. Some interesting aspects are found when strongly ineracting, elonganted, or functional inorganic nanoparticles are introduced.

After the initial works of Haraguchi et al. [101], a variety of works can be found on the incorporation of partly or fully exfoliated clay nanosheets into hydrogels. The hybrid

hydrogels are prepared by in situ radical polymerization using a specific solution system. Highly stable, structurally homogeneous materials with extraordinary mechanical properties are achieved. Static Light Scattering (SLS) and rheology experiments on exfoliated laponite-containing gels indicate that the cross-linking proceeds mainly through the nanosheets, even if the exact nature is still subject of discussions. In the meantime, similar particles have been employed analogously in the combination with thermoresponsive PNiPAAm hydrogels [102]. The gels have been tensile-tested under aqueous conditions at various temperatures. The response time of the materials could further be improved by preparing porous materials by the salt-leaching technique [103], or by the incorporation of carboxymethyldextran [104].

Further functionalization result in additional incorporation of pH-responsive components [105]. The possibility to manipulate the position, orientation and dynamics of ferro- and superparamagnetic nanoparticles by the means of extermal magnetic fields is of interest for a range of scientific and technological applucations. Among the latter is the use of magnetic fluids (dispersions of magnetic nanoparticles in a carrier fluid) in vacuum sealing rotary feedthroughs, e.g., in hard disk drives, and in high quality loudspeaker membranes. In both applications, the property of a viscous fluid (lubrication, and as damping and cooling liquid) is needed combined with the possibility to locate the fluid in space by a permanent magnet. Additionally, magnetic colloids are in use in dampers and separators making use of their field-dependent properties.

Actual investigations also concern nanomotors and nanoactuators. An intersting concept is to use the field-induced translational or rotational motion of tracer particles to evaluate the dynamic behavior of soft materials. Such techniques are known as nano- or microrheology, offering advantages like small sample volume, an extended frequency range in comparison with macroscopic approaches, and the option to extract spatially resolved information.
In magnetic particle nanorheology, the probe is actively driven within the material, either in oscillatory or steady motion. It was shown for different systems that the rotational activation of magnetically blocked micro- or nanoscaled particles in an oscillatory field can be used to investigate the viscosity of (non-)Newtonian fluids at different temperatures [107]
Gels filled with magnetic nanoparticles are commonly known as ferrogels. After pioneering work of Zrınyi [115] demonstrating that it is possible to use the susceptibility of superparamagnetic particles, and the induction of particle interaction in static magnetic fields to achieve considerable gel bending or contraction effects. Meanwhile, magnetostrictive gels based on different matrices like PVA [116], Pluronics [117], and PHEMA have been realized.

If the synthesis of magnetically loaded gels and elastomers is performed under the impact of an external magnetic field, magnetically and mechanically anisotropic materials are obtained with an intrinsic magnetic anisotropy, manifesting itself both as direction dependent elastic modulus and direction dependent swelling. The mechanical properties are further influenced by external fields. It is assumed that next to particle-field and particle-particle interactions, chain formation plays a significant role for the stiffening effect.

Therapeutic application potential is found in Magnetic Drug Targeting [118], exploiting the magnetic moment of the particles for a preferred enrichment at the center of disease, e.g., a tumor, by the application of an outer permanent magnet. A predominantly local release of drugs like cytostatica is achieved, resulting in a decrease in side effects.

4.2 Optical Detection & Diagnostics

The size-selective interaction of functional nanoparticles with electromagnetic fields is of considerable interest for their application as optical probes and in diagnostics.
latest developments in biocompatible functionalized nanoparticles show considerable promise for both enhanced and novel applications in the biomedical and diagnostic fields, ranging from targeted drug delivery to contrast enhancement in imaging applications. The optical properties of noble metal nanoparticles lead to many uses as sensing and imaging techniques. Mirkin and co-workers [119] have pioneered the use of DNA in assembling noble metal particles, and studying their application for the colorimetric detection of biological targets by optical means. The use of nanoparticles in the field of photonics is immense but beyond the scope of this lecture.

The optical detection of the particle location and state of agglometration is possible by means of absorption, fluorescence of the particles itself or bound organic moeieties, or scattering.
The sensitivity of the surface plasmon resonance absorption of the particles to the immediate particle environment enables the detection of adsorption or binding processes taking place on the specifically designed particle surface. The nanoparticles can be functionalized to observe only the molecules of interest and the absorption of desired molecules can be observed by a shift in the plasmon resonance absorption (i.e. a change in the color). Haes et al. [120] have used the surface plasmon resonance from an array of silver nanoparticles created by nanosphere lithography to detect the interaction of amyloid β-derived diffusible ligands (ADDL) and anti-ADDL antibody, believed to be important in Alzheimer's disease. The binding constant of the anti-ADDL and ADDL can be determined with this technique.
The localized surface plasmon resonance effect is also responsile with the strong impact of metal nanoparticles on the fluorescence of chromophores in the vicinity of the particle surface. Due to the strong electromagnetic field gradients generated at the surface of metal nanoparticles under irradiation [121], the fluorescence of chromophores located within ~5 nm of the surface is quenched, while that of chromophores at distances of ~10 nm or larger is strongly enhanced.

Further, the enhanced scattering cross section due to surface field effects can be used as a powerful technique to image biological systems. Gold nanoparticle surface plasmon resonance scattering is predicted in the Mie equations and is found to increase as the size of the nanoparticle increases. By conjugating gold nanoparticles to anti-EGFR antibody, it is possible to distinguish between cancer and non-cancer cells from the strong scattering images of the gold nanoparticles conjugated to antibodies that binds only to the cancer, but not to the non-cancer cells [123].
The surface-enhanced raman scattering (SERS) of molecules deposited on rough noble metal surfaces and particle aggregates was observed in the early 1970s [124] and is now interpreted to be originated by a “hot spot” effect at the junction of two or more particles [125]. As a method that combines a large signal enhancement along with chemical information, SERS is being developed as a powerful diagnostic tool [125].
Similarly, the interaction of magnetic nanoparticles with suitable magnetic fields can be used for detection and imaging applications. In vitro applications aim on the detection (magnetorelaxometry assays) or the separation (magnetic cell separation) [126] of biological species like proteins [127], oligonucleotides [129], or cells. Magnetic nanoparticles are now routinely used in MRI nedical applications to enhance the contrast between biological structures and for early tumor detection. The application as a contrast-enhancing agent in magnetic resonance imaging is based on the effect that the magnetic field induced by a

particle results in modified relaxation times in the surrounding tissue. In addition, magnetic nanoparticles can replace the radioactive materials that are currently used as drug tracers. The detection and quantification of the drug location is then possible by measuring magnetic instead of radioactivity intensities, eliminating potential human harm from radiation.

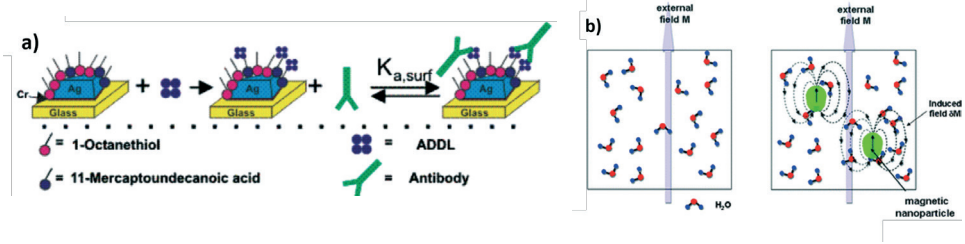


Fig. 5: *a) Design of the LSPR biosensor for anti-ADDL detection. Reprinted with permission from ACS Nano Lett.*[120]. *b) Magnetic contrast effect of magnetic nanoparticles in water. The induced magnetic field of the magnetic nanoparticles perturbs the magnetic relaxation processes of the protons in water molecules, resulting in T_2 shortening of the proton with a dark MR contrast. Reprinted with permission from RCS, Chem. Commun. 2007* [91].

4.3 Local heating

The local heat development by functional nanoparticles is possible under resonant conditions as well from the localized surface plasmon resonance as under magnetic resonance for accordingly designed particles. Under ideal conditions, the particles show a high efficiency of heat development and high selectivity. While the LSPR signal is very intense and show a narrow dispersion, the resonant frequencies in the visible / UV region limit the penetration depth within real applied systems due to the absorption by the particle environment. One advantage of the plasmonic systems is the possibility to spectroscopically and time-resolved read out the actual particle temperature during the heating process with high accuracy.
In the case of magnetic particles, the magnetic resonance can be tuned to occur in the kHz range, with the advantage that many other materials incuding biological tissue are largely transparent in this range, allowing a highly localized effect, by cost of efficiency.
In the biomedical field, magnetic nanoparticles and their dispersions play an important role for the development of new diagnostics and therapeutics. Due to the limited stability and compatibility of metallic particles, in general oxidic ferrites are used, predominantly magnetite (Fe_3O_4) or maghemite (γ-Fe_2O_3) with their accepted biocompatibility. Hyperthermia is a therapeutic procedure to increase the temperature of a certain body region to decelerate the growth of tumor tissue, respectively to sensitize, damage or ideally destroy them. For magnetic fluid hypethermia [53], superparamagnetic particles are injected into the tumor, with the advantage that the heat development can be limited to the tumor region by magnetic heating of the particles in an applied AC field.

The magnetic heatability of magnetic nanoparticles is furthermore exploited for glue formulations that can be activated by AC magnetic fields [131].
The combination of magnetoresponsive and thermoresponsive properties is of interest from different points of view. The thermal properties can be used to complement or to control the

features of magnetic fluids, or the physical properties of the nanomagnets can add the performance of thermoresponsive polymer systems.

4.4 Multifunctional

Multicomponent hybrid nanoparticles demonstrate many interesting physical properties originating from intercomponent interactions. Core/shell or dumbbell nanoparticles are attractive multifunctional systems for potential applications ranging from catalysis, advanced biomedicine, data storage, and high-frequency materials to advanced permanent magnets [133].
Biocompatible Au@Fe_3O_4 nanoparticles that selectively attach to A431 human epithelial carcinoma cell line [137] are magnetically and optically active and are useful for simultaneous magnetic and optical detection. Furthermore, an external magnetic field can be applied to manipulate the cells.

Core/shell nanoparticles with a magnetic core and a non-magnetic shell, such as FePt/SiO2 and FePt/MnO, have been developed by several groups as building blocks for self-organized magnetic media [138], where FePt with high anisotropy provides the high thermal stability, and a non-magnetic shell can be used to decouple the grains and prevent agglomeration of the nanoparticles. A soft magnetic-dielectric core/shell nanoparticle may have applications in radio-frequency (RF) transducer materials, where the dielectric shell can be used to cut eddy current loss [140]. In the case of hard-soft exchange-coupled magnetic core/shell nanoparticles, both the coercivity and magnetization can be tuned by material parameters and the dimensions of the core and shell, as well as their mutual interactions [141].

Multicomponent magnetic-semiconductor nanoparticles show potential for applications in spintronics. The magnetic component will provide spin filtering of the carriers by injection or extraction; and the recombination of the electrons and holes in the semiconductor component is expected to give polarized emission. The structure can be considered as a nano spin-LED, except that the device is directly synthesized from the solution phase instead of fabricated by top-down lithography techniques [75].

References

[1] A. N. Shipway, and I. Willner, Chem. Commun. 2035 (2001)
[2] B. L. Cushing, V. L. Kolesnichenko, C. J. O'Connor, Chem. Rev. 104, 3893 (2004)
[3] D. Sander, Curr. Opin. Solid State Mater. Sci. 7, 51 (2003)
[4] C. R. Berry, Phys. Rev. 88, 596 (1952)
[5] T. de Planta, R. Ghez, and F. Piuz, Helv. Phys. Acta 37, 74 (1964)
[6] H. J. Wasserman and J. S. Vermaak, Surf. Sci. 22, 164 (1970)
[7] A. N. Goldstein, C. M. Echer, and A. P. Alivisatos, Science 256, 1425 (1992)
[8] R. Dingreville, J. Qu, and M. Cherkaoui, J. Mech. Phys. Solids, 53 1827 (2005)
[9] H. Duan, J. Wang, Z. Huang, and B. Karihaloo, J. Mech. Phys. Solids, 53 1574 (2005)
[10] P. Sharma, and S. Ganti, J. Appl. Mech. 71 663 (2004)
[11] G. Kickelbick (ed.), Hybrid Materials (Wiley, Weinheim, 2006)
[12] W. Loos and F. Du Prez, Macromol. Symp., 210, 483 (2004)
[13] V. G. Ngo, C. Bressy, C. Leroux and A. Margaillan, Polymer, 50, 3095 (2009)
[14] N. Frickel, R. Messing, T. Gelbrich and A. M. Schmidt, Langmuir, 26, 2839 (2010)
[15] H. B. Sunkara, J. M. Jethmalani and W. T. Ford, Chem. Mater. 6, 362 (1994)
[16] E. Bourgeat-Lami, M. Insulaire, S. Reculusa, A. Perro, S. Ravaine and E. Duguet, J. Nanosci. Nanotechnol. 6, 432 (2006)
[17] S. Eustis, and M. El-Sayed, Chem. Soc. Rev. 35, 209 (2006)
[18] E. Petryayeva, and U. J. Krull, Anal Chim Acta 706, 8, (2011)
[19] S. A. Maier, and H. A. Atwater, J. Appl. Phys. 98, 011101 (2005)
[20] K. L. Kelly, E. Coronado, L. L. Zhao, and G. C. Schatz, J. Phys. Chem. B 107, 668 (2003)
[21] G. A. Ozin, A. C. Arsenault, and L. Cademartiri, Nanochemistry: A Chemical Approach to Nanomaterials, RSC publishers, Cambridge, 2009.
[22] J. Perez-Juste, I. Pastoriza-Santos, L. M. Liz-Marzan, and P. Mulvaney, Coord. Chem. Rev. 249, 1870 (2005)
[23] S. Link, and M.A. El-Sayed, J. Phys. Chem. B 103, 8410 (1999)
[24] M. Hu, J.Y. Chen, Z.Y. Li, L. Au, G.V. Hartland, X.D. Li, M. Marquez, and Y.N. Xia, Chem. Soc. Rev. 35, 1084 (2006)
[25] J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao, and R. P. Van Duyne, Nat. Mater. 7, 442 (2008)
[26] K. L. Kelly, E. Coronado, L. L. Zhao and G. C. Schatz, J. Phys. Chem. B 107, 668 (2003)
[27] M. A. El-Sayed, Acc. Chem. Res. 34, 257 (2001)
[28] M. P. Pileni, Nat. Mater. 2, 145 (2003)
[29] R. C. Jin, Y. W. Cao, C. A. Mirkin, K. L. Kelly, G. C. Schatz and J. G. Zheng, Science, 294, 1901 (2001)
[30] K.A. Willets, R.P. Van Duyne, Annu. Rev. Phys. Chem. 58, 267 (2007)
[31] L.M. Liz-Marzan, Langmuir 22, 32 (2006)
[32] K. S. Lee, and M. A. El-Sayed, J. Phys. Chem. B. 110, 19220 (2006)
[33] Y. Choi, Y. Park, T. Kang, and L. P. Lee, Nat. Nanotechnol. 4, 742 (2009)
[34] J. K. Jaiswal, H. Mattoussi, J. M. Mauro, and S. M. Simon, Nat. Biotechnol 21, 47 (2003)
[35] M. F. Frasco, and N. Chaniotakis, Sensors 9, 7266 (2009)
[36] X. M. Lu, M. Rycenga, S. E. Skrabalak, B. Wiley, and Y. N. Xia, Annu. Rev. Phys. Chem. 60, 167 (2009)
[37] J. J. Mock, M. Barbic, D. R. Smith, D. A. Schultz, and S. J. Schultz, Chem. Phys. 116, 6755 (2002)

[38] A. L. Efros, Sov. Phys. Semicond. 16, 772 (1982)
[39] A. I. Ekimov, A. A. Onushenko, Sov. Phys. Semicond. 16, 775 (1982)
[40] A. Henglein, J. Phys. Chern. 86, 2291 (1982)
[41] R. Rossetti, S. Nakahara, and L. E. Brus, J. Chem. Phys. 79, 1086 (1983)
[42] M. C. Schlamp, X. Peng, and A.P. Alivisatos, J. Appl. Phys. 82, 5837 (1997)
[43] H. Matioussi, L. H. Radzilowski, B. O. Dabboussi, E. L. Thomas, and M. F. Rubner, J. Appl. Phys. 83, 7965 (1998)
[44] M. Y. Gao, C. Lesser, S. Kirstein, H. Möhwald, A L. Rogach, and H. Weller, J. Appl. Phys. 87, 2297 (2000)
[45] M. Han, X. Gao, J.Z. Su, and S. Nie, Nature Biotechn. 19, 631 (2001)
[46] L. E. Brus, J. Chern. Phys. 80, 4403 (1984)
[47] D. L. Leslie-Peleckie, and R. D. Rieke, Chem. Mater. 8, 1770 (1996)
[48] I. S. Jacobs, and C. P. Bean, in: Rado GT, Suhl H (eds): Magnetism, Vol. III (Academic, New York, 1963)
[49] E. C. Stoner, and F. R. S. Wohlfarth, Philos. Trans. R. Soc. Lond A 240, 599 (1948)
[50] L. Néel, Ann Geophys 5, 99 (1949)
[51] W. F. Brown Jr., J. Appl. Phys. Suppl. 30, 130S (1959)
[52] L. Néel, in: C. J. Porter (ed): Progress in Low Temperature Physics, Vol 1 (North Holland Publisher Cooperation, Amsterdam, 1964)
[53] G. Glöckl, R. Hergt, M. Zeisberger, S. Dutz, S. Nagel, and W. Weitschies, J. Phys. Condens. Matter 18, S2935 (2006)
[54] M. T. Thompson, IEEE Trans. Magn. 34, 3755 (1998)
[55] X. K. Zhang, Y. F. Li, and J. Q. Xiao, J. Appl. Phys. 93, 7124 (2003)
[56] I. A. Brezovich, Med. Phys. Monogr. 16, 82 (1988)
[57] F. Söffge, and E. Schmidtbauer, J. Magn. Magn. Mater. 24, 54 (1981)
[58] S. Dutz, R. Hergt, J. Mürbe, R. Müller, M. Zeisberger, W. Andrä, J. Töpfer, and M. E. Bellemann, J. Magn. Magn. Mater. 308, 305 (2007)
[59] H. Bönnemann, and R. M. Richards, Eur. J. Inorg. Chem. 2001, 2455 (2001)
[60] E. Gaffet, M. Tachikart, O. El Kedim, and R. Rahouadj, Mater. Charact. 36, 185 (1996)
[61] A. Amulyavichus, A. Daugvila, R. Davidonis, and C. Sipavichus, Fiz. Met. Metalloved. 85, 111 (1998)
[62] A. Schalnikoff, and R. Roginsky, Kolloid Z. 43, 67 (1927)
[63] J. R. Blackborrow, D. Young, Metal Vapor Synthesis (Springer Verlag, New York, 1979)
[64] K. J. Klabunde, Y.-X. Li, and B.-J. Tan, Chem. Mater. 3, 30 (1991)
[65] J. S. Bradley, in: G. Schmid (ed): Clusters and Colloids (VCH, Weinheim, 1994), p. 477f.
[66] K. J. Klabunde, and G. C. Cardenas-Trivino, in: A. Fürstner (ed): Active Metals (VCH, Weinheim, 1996), p. 237ff
[67] G. Schmid, M. Bäumle, and N. Beyer, Angew. Chem. Int. Ed. 39, 182 (2000)
[68] H. Cölfen, and T. Pauck, Colloid Polym. Sci. 275, 175 (1997)
[69] M. Faraday, Philos. Trans. R. Soc. London 147, 145 (1857)
[70] J. Turkevich, P. C. Stevenson, and J. Hillier, Disc. Faraday Soc. 11, 55 (1951)
[71] J. Turkevich, and G. Kim, Science 169, 873 (1970)
[72] J. Turkevich, Gold Bull. 18, 86 (2985)
[73] V. K. LaMer and R. H. Dinegar, J. Am. Chem. Soc. 72, 4847 (1950)
[74] Y. Xia, Y. Xiong, B. Lim, and S. E. Skrabalak, Angew. Chem. Int. Ed. 48, 60 (2009)
[75] S. Sun, and H. Zeng, Adv. Funct. Mater. 18, 391 (2008)
[76] J. D. Aiken, and R. G. Finke, J. Mol. Catal. A 145, 1 (1999)
[77] R. Jin, S. Egusa, and N. F. Scherer, J. Am. Chem. Soc. 126, 9900 (2004)
[78] H. Kim, M. Achermann, L. P. Balet, J. A. Hollingsworth, V. I. Klimov, J. Am. Chem.

Soc. 2005, 127, 544
[79] H. Zeng, J. Li, Z. L. Wang, J. P. Liu, and S. H. Sun, Nano Lett. 4, 187 (2004)
[80] S. J. Oldenburg, R. D. Averitt, S. L. Westcott, and N. J. Halas, Chem. Phys. Lett. 288, 243 (1998)
[81] W. L. Shi, Y. Sahoo, M. T. Swihart, and P. N. Prasad, Langmuir 21, 1610 (2005)
[82] A. R. Kortan, R. Hull, R. L. Opila, M. G. Bawendi, M. L. Steigerwald, P. J. Carroll, and L. E. Brus, J. Am. Chem. Soc. 112, 1327 (1990)
[83] H. Yu, M. Chen, P. M. Rice, S. X. Wang, R. L. White, and S. H. Sun, Nano Lett. 5, 379 (2005)
[84] H. W. Gu, R. K. Zheng, X. X. Zhang, and B. Xu, J. Am. Chem. Soc. 126, 5664 (2004)
[85] H. W. Gu, Z. M. Yang, J. H. Gao, C. K. Chang, and B. Xu, J. Am. Chem. Soc. 127, 34 (2005)
[86] K. W. Kwon, and M. Shim, J. Am. Chem. Soc. 127, 10269 (2005)
[87] S. Kudera, L. Carbone, M. F. Casula, R. Cingolani, A. Falqui, E. Snoeck, W. J. Parak, and L. Manna, Nano Lett. 5, 445 (2005)
[88] T. Mokari, E. Rothenberg, I. Popov, R. Costi, and U. Banin, Science 304, 1787 (2004)
[89] T. Mokari, C. G. Sztrum, A. Salant, E. Rabani, and U. Banin, Nat. Mater. 4, 855 (2005)
[90] D. C. Lee, D. K. Smith, A. T. Heitsch, and B. A. Korgel, Annu. Rep. Prog. Chem. Sect. C Phys. Chem. 103, 351 (2007)
[91] Y. W. Jun, J. S. Choi, and J. Cheon, Chem. Commun. 1203 (2007)
[92] P. D. Cozzoli, T. Pellegrino, and L. Manna, Chem. Soc. Rev. 35, 1195 (2006)
[93] D. Myers, Surfaces, Interfaces, and Colloids (2nd edn., Wiley-VCH, New York,1999)
[94] V. S. Zaitsev, D. S. Filimonov, I. A. Presnyakov, R. J. Gambino, and B. Chu, J. Colloid Interface Sci. 212, 49 (1999)
[95] J. Lee, M. Senna, Colloid Polym. Sci. 273, 76 (1995)
[96] A. Kondo, H. Fukuda Colloids Surf. A Physicochem. Eng. Asp. 153, 43 (1995)
[97] A. Bansal, H. C. Yang, C. Z. Li, K. W. Cho, B. C. Benicewicz, S. K. Kumar and L. S. Schadler, Nat. Mater. 4, 693 (2005)
[98] D. Ciprari, K. Jacob, and R. Tannenbaum, Macromolecules 39, 6565 (2006)
[99] G. Tsagaropoulos, and A. Eisenberg, Macromolecules 28, 6067 (1995)
[100] D. Cangialosi, V. M. Boucher, A. Alegria, and J. Colmenero, Soft Matter 9, 8619 (2013)
[101] K. Haraguchi and T. Takehisa, Adv. Mater. 14, 1120 (2002)
[102] K. Haraguchi, S. Taniguchi and T. Takehisa, ChemPhysChem, 6, 238 (2005)
[103] J. H. Ma, L. Zhang, Z. Li and B. R. Liang, Polym. Bull. 61,593 (2008)
[104] J. H. Ma, L. Zhang, B. Fan, Y. J. Xu and B. R. Liang, J. Polym. Sci. Part B Polym. Phys. 46, 1546–1555 (2008)
[105] J. H. Ma, Y. J. Xu, Q. S. Zhang, L. S. Zha and B. R. Liang, Colloid Polym. Sci. 285, 479–484 (2006)
[106] X. B. Hu, L. J. Xiong, T. Wang, Z. M. Lin, X. X. Liu and Z. Tong, Polymer 50, 1933 (2006)
[107] J. Wagner, B. Fischer, T. Autenrieth, R. Hempelmann, J. Phys. Condens. Matter 18, S2697 (2006)
[108] C. Barrera, V. Florián-Algarin, A. Acevedo, C. Rinaldi, Soft Matter 6, 3662 (2010)
[109] V. L. Calero-Diaz del Castillo, D. I. Santiago-Quiñonez, C. Rinaldi, Soft Matter 7, 4497 (2011)
[110] F. Ludwig, A. Guillaume, M. Schilling, N. Frickel, and A. M. Schmidt, J. Appl. Phys. 108, 033918 (2010)
[111] A. M. Schmidt, Colloid Polym. Sci. 285, 953 – 966 (2007)
[112] M. Feyen, E. Heim, F. Ludwig, and A. M. Schmidt, Chem Mater 20, 2942 (2008)
[113] A. Kaiser, and A. M. Schmidt, J. Phys. Chem. B 112, 1894 (2008)

[114] E. Roeben, L. Roeder, S. Teusch, M. Effertz, U. K. Deiters, and A. M. Schmidt, Colloid Polym. Sci. 140, 199-216 (2014)
[115] M. Zrınyi, Colloid Polym. Sci. 278, 98 (2000)
[116] R. V. Ramanujan and L. L. Lao, Smart Mater. Struct. 15, 952 (2006)
[117] J. Qin and J. Qin, Adv. Mater. 21, 1354 (2009)
[118] C. Alexiou, W. Arnold, R. J. Klein, F. G. Parak, P. Hulin, C. Bergemann, W. Erhardt, S. Wagenpfeil, A. S. Lübbe, Cancer Res 60, 6641 (2000)
[119] N. L. Rosi and C. A. Mirkin, Chem. Rev. 105, 1547 (2005)
[120] A. J. Haes, W. P. Hall, L. Chang, W. L. Klein and R. P. Van Duyne, Nano Lett. 4, 1029 (2004)
[121] J. Lakowicz, C. Geddes, I. Gryczynski, J. Malicka, Z. Gryczynski, K. Aslan, J. Lukomska, E. Matveeva, J. Zhang, R. Badugu and J. Huang, J. Fluoresc. 14, 425 (2004)
[122] K. G. Thomas and P. V. Kamat, Acc. Chem. Res. 36, 888 (2003)
[123] I. H. El-Sayed, X. Huang and M. A. El-Sayed, Nano Lett. 5, 829 (2005)
[124] A. Campion and P. Kambhampati, Chem. Soc. Rev. 27, 241 (1998)
[125] S. Nie and S. R. Emory, Science 275, 1102 (1997)
[126] I. Safarik, and M. Safarikova, J. Chromatogr. B 722, 33 (1999)
[127] M. A. Burns, G. I. Kvesitadze, and D. J. Graves, Biotechnol Bioeng 27, 137 (2004)
[128] G. K. Kouassi, J. Iruyadaraj, and G. McCarthy, J. Nanobiotechnol 3, 1 (2005)
[129] J. Krizova, A. Spanova, B. Rittich, and D. Horak, J. Chromatogr. A 1064 (2005)
[130] D. Horak, B. Rittich, J. Safar, A. Spanova, J. Lenfeld, and M. J. Beneš, Biotechnol. Progr. 17, 447 (2001)
[131] C. Kirsten, M. Roth, O. Lammerschop, and H.-M. Sauer, Patent WO 02/13580 A1 (2002)
[132] P. Stark, G. Rossi, H. R. Mojazza, C. A. R. Haghighat, and P. Schuler, Patent PCT/US01/14079
[133] H. W. Gu, Z. M. Yang, J. H. Gao, C. K. Chang, and B. Xu, J. Am. Chem. Soc. 127, 34 (2005)
[134] J. S. Choi, Y. W. Jun, S. I. Yeon, H. C. Kim, J. S. Shin, and J. Cheon, J. Am. Chem. Soc. 128, 15982 (2006).
[135] S. T. Selvan, P. K. Patra, C. Y. Ang, and J. Y. Ying, Angew. Chem. Int. Ed. 46, 2448 (2007).
[136] Y. S. Lin, S. H. Wu, Y. Hung, Y. H. Chou, C. Chang, M. L. Lin, C. P. Tsai, and C. Y. Mou, Chem. Mater. 18, 5170 (2006)
[137] C. Xu, J. Xie, D. Ho, C. Wang, N. Kohler, E. G. Walsh, J. R. Morgan, Y. E. Chin, and S. Sun, Angew. Chem. Int. Ed. 47, 173 (2008)
[138] C. H. Jun, Y. J. Park, Y. R. Yeon, J. R. Choi, W. R. Lee, S. J. Ko, and J. Cheon, Chem. Commun. 1619 (2006)
[139] V. Salgueirino-Maceira, M. A. Correa-Duarte, and M. Farle, Small 1, 1073 (2005)
[140] J. Li, H. Zeng, S. H. Sun, J. P. Liu, and Z. L. Wang, J. Phys. Chem. B 108, 14005 (2004)
[141] H. Zeng, J. Li, Z. L. Wang, J. P. Liu, and S. H. Sun, Nano Lett. 4, 187 (2004)

A 2 Biomolecules

A. Baumann
Institute of Complex Systems, ICS-4
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Cells, the factories of biomolecules **2**

2.1 Structural features of eukaryotic cells **4**

3 Biomolecules **6**

3.1 Molecular Composition of Cellular Membranes **6**

3.2 The Nucleus **9**

3.2.1 Nucleic Acids **9**

3.2.2 Synthesis of proteins **12**

3.2.3 Cell regulation by epigenetic factors **13**

4 Summary **15**

References **16**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

The term 'biomolecule' covers any molecule that is produced by a living organism. This includes small molecules such as hormones, metabolites, neurotransmitters, vitamins as well as amino acids, fatty acids, lipids, or saccharides. From these small molecules (large) macromolecules like nucleic acids, polysaccharides, or proteins can be build. The capability to produce small and large biomolecules is an inherent property of the elementary and functional units of each living organism: the cell. In addition to produce biomolecules, cells provide an organism with a variety of fascinating features. They assemble to larger entities and form different tissues, they are able to self-replicate, i. e. divide into two identical daughter cells, and cells equip an organism with efficient communication systems. To fulfil these challenging tasks cells come in remarkably diverse structures and functions within a single organism. However, all cells share fundamental biochemical and biophysical principles. In the following, I will briefly introduce the main cellular 'blueprint' before I am going to focus on the properties of some essential biomolecules that are produced in a living cell. This is a tutorial for physicists interested in biochemical and molecular cell function. There are several superb textbooks available that comprehensively cover this research area. The text and figures of this review are, in part, based on the text books "Biochemistry" by Berg et al. [1], "Lehninger, Principles of Biochemistry" by Nelson and Cox [2], "Molecular Cell Biology" by Lodish et al. [3], and "Molecular Biology of the Cell" by Alberts et al. [4]. The interested reader is also referred to other textbooks cited at the end of the text [5 - 7].

2 Cells, the factories of biomolecules

Multicellular organisms contain many different types of cells, which vary in size, shape, and function. The human body, for example, consists of about 200 cell types that collectively amount to more than 10^{14} cells. Typically eukaryotic cells are 5 to 100 μm in diameter and thus much larger than e.g. bacteria (length ~2 μm, diameter ~1 μm). To provide some insight into their functional divergence I will briefly introduce a few eukaryotic cell types: neurons, erythrocytes, and hepatocytes.

The human brain consists of approximately 10^{11} *neurons*. Both, in the central and the peripheral nervous system, neurons equip the organism with an efficient system for fast communication. Some cells of the nervous system (*motor neurons*) possess very long cellular extensions (Fig. 1). These *axons* can be ~1 m long. They originate in the spinal cord and pass electrical signals to muscle fibres in legs or toes.

To convey information to target cells, neurons generate *action potentials.* These are small electrical discharges of the membrane potential that originate from the consecutive activity of ion channel proteins housed in the cell membrane (see also chapter A5). The action potential travels down the axon with velocities of up to 100 m/sec. When the action potential reaches the *presynaptic terminal* region of the neuron, it stimulates the release of chemical *transmitters*. The transmitters diffuse to a target cell and activate specific proteins that eventually generate an electrical response of that cell.

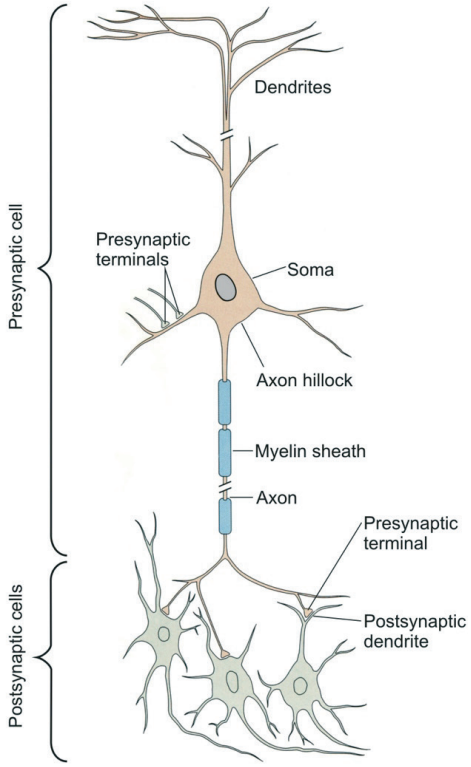


Fig. 1: *Schematic drawing of a neuron. The dendrites serve as the input region of a neuron and receive information from other neurons or the environment. The soma houses the nucleus with the DNA and also is the place where new proteins are synthesized. The soma gives rise to the axon that propagates information to target cells. Axon lengths can range from µm to m scales. Modified from [7].*

Cell types contained in the blood fulfil functions ranging from the transport of oxygen to the production of antibodies. They all have limited life spans and are continuously produced throughout an animal's life. 'White' blood cells, *leukocytes*, combat infection and engulf and destroy debris. 'Red' blood cells, *erythrocytes* (Fig. 2), are specialized for carrying O_2 and CO_2. Human erythrocytes are small, biconcave disks with a diameter of 6 to 9 µm. During their maturation, erythrocytes produce large amounts of haemoglobin (up to 34% of the cells total weight).

Functional haemoglobin assembles from four subunits. Each subunit contains a prosthetic group called *heme*. The heme group is a porphyrin system coordinating a single Fe^{2+} atom to which O_2 can bind. Haemoglobin is saturated with O_2 in the lung and from there circulating erythrocytes provide O_2 supply of peripheral tissue via the blood stream.

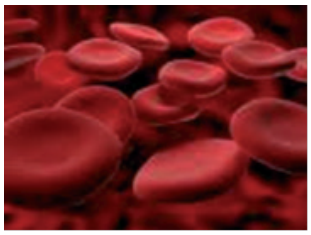


Fig. 2: *Erythrocytes are filled with haemoglobin and provide oxygen supply in the body.*

The liver is the largest gland of the human body. Cells in the liver, *hepatocytes* (Fig. 3), are arranged in folded sheets and face blood-filled spaces called sinusoids. The blood is separated from the surface of the hepatocytes by a single layer of epithelial cells covering both sides of each hepatocyte sheet. This arrangement facilitates the exchange of metabolites between hepatocytes and the blood.



Fig. 3: *Hepatocytes synthesize, store, and degrade a huge number of different substances in the liver.*

The liver is the key organ where nutrients that have been absorbed from the gut and then transferred to the blood are processed for further use in other cells of the body. Hepatocytes thus participate in the synthesis, degradation, and storage of a huge number of different substances. Hepatocytes differ in their life-style, e. g. from cells that line the lumen of the gut. The latter are facing very harsh chemical conditions originating from the corrosive contents of the gut. Consequently they are rapidly and continuously replaced by newborn cells. Hepatocytes are protected from direct interaction with the gut contents. Therefore, they do not undergo a rapid turnover and are replaced at a slow but precisely controlled rate. However, if two-thirds of a rat's liver is surgically removed, a liver of normal size can regenerate from the remaining tissue within a week.

2.1 Structural features of eukaryotic cells

Despite their morphological and functional differences, all eukaryotic cells share important features that manifest in conserved subcellular structures (Fig. 4). The *plasma membrane* defines the surface of the cell, separating its contents from the surrounding. The external surface of a cell is in contact with other cells and the extracellular fluid harbouring solutes and nutrient molecules. The plasma membrane is held together primarily by non-covalent hydrophobic interactions and forms a thin, hydrophobic layer around the cell. The membrane is a barrier to the free passage of inorganic ions and most other charged or polar compounds. Nevertheless, cells continuously have to exchange "substances" with their environment. To manage this logistic problem, specific transport proteins in the plasma membrane allow the

passage of certain ions and molecules. Other membrane proteins not participating in 'cargo' processes include receptors that transmit signals from the outside to the inside of the cell and enzymes that participate in membrane-associated reaction pathways. In addition to the plasma membrane, eukaryotic cells contain internal membranes that form the boundaries of organelles such as the endoplasmic reticulum, Golgi complex, lysosomes, mitochondria, and nucleus.

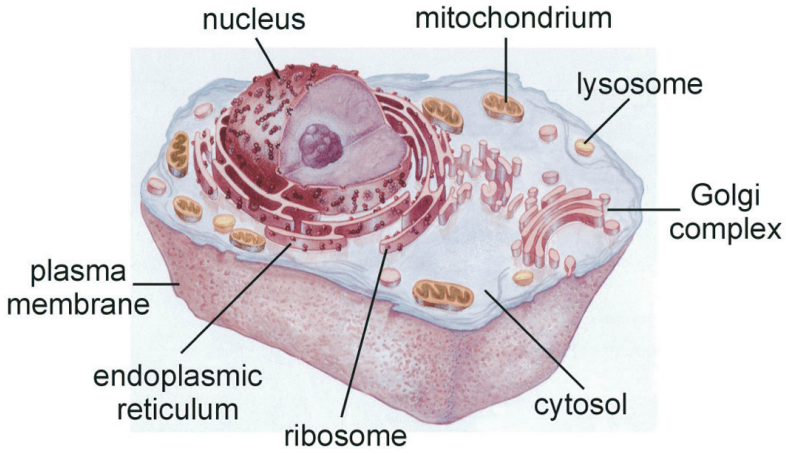


Fig. 4: *Schematic drawing of a typical animal cell. Animal cells are usually 2 to 30 µm in diameter. Structures and subcellular components are labelled and explained in the text. Modified from [2].*

The internal volume bounded by the plasma membrane, the *cytosol*, is composed of an aqueous solution and a variety of insoluble, suspended particles. The cytosol is a highly concentrated aqueous solution with complex composition and gel-like consistency. A major constituent of the cytosol is the *cytoskeleton*, a meshwork-like structure composed of polymeric protein fibres, called microtubules and filaments that participate in maintaining a cell's shape and mobility and also provide anchoring points for other cellular structures. Suspended in the cytosol are many proteins, small organic molecules (metabolites), and intermediates of biosynthetic and degradative pathways. Among the particles operating in the cytosol are the *ribosomes*. They are the sites at which protein synthesis occurs (see also chapter E2). Ribosomes engaged in protein synthesis often come in clusters called polysomes and are located on the outer surface of the *endoplasmic reticulum*.

The endoplasmic reticulum (ER) is a highly convoluted, three dimensional network of membrane-enclosed spaces extending throughout the cytoplasm and enclosing a subcellular compartment separate from the cytoplasm. The flattened branches of this compartment are continuous with each other and with the nuclear envelope. Ribosomes that synthesize proteins destined for plasma membrane insertion or export attach to the surface of the ER, and these proteins are passed through the membrane into the ER lumen as they are synthesized. By contrast, proteins remaining in the cytosol are synthesized on ribosomes unassociated with the ER. Areas where thousands of ribosomes are attached to the ER are called *rough endoplasmic reticulum*. Although physically continuous with the rough ER, the smooth ER is free of ribosomes. Its main function is in lipid biosynthesis.

Most eukaryotic cells have *Golgi complexes* (see Fig. 4). Structurally and functionally the Golgi complex is asymmetric. The cis side faces the rough ER, and the trans side faces the plasma membrane. Newly synthesized secretory and membrane proteins move into the lumen of the rough ER and pass through the Golgi complex to the trans side. In the Golgi complex, proteins become modified by sulfate groups, carbohydrates, or lipid moieties. These modifications serve as molecular "labels" directing a protein to its final destination properly.

Lysosomes (see Fig. 4) are spherical vesicles usually 1 μm in diameter. They contain enzymes capable of digesting proteins, polysaccharides, nucleic acids, and lipids. Lysosomes function as cellular recycling centres. They contain degradative enzymes which break down complex molecules, for example nutrients, but also worn-out organelles from the cell's own cytoplasm. The enzymes in principle could act on all cellular components were they not confined by the lysosomal membrane. Another strategy to keeping these enzymes inactive in the cytoplasm is their exquisite pH-dependence of activity which is maximal in the lysosomal lumen ($pH \leq 5$) compared to the cytosol ($pH \approx 7$).

A pivotal organelle of all eukaryotic cells is the *mitochondrion*. Mitochondria typically have a diameter of 1 μm. Mitochondrial enzymes catalyze the oxidation of organic nutrients by molecular oxygen. The chemical energy released in mitochondrial oxidations is used to generate adenosine triphosphate (ATP), the major energy-carrying molecule of cells. When glucose is completely oxidized to CO_2 and H_2O, 26 molecules of ATP are produced. The ATP diffuses to all parts of the cell and provides the fuel for cellular work. This energy is supplied by the hydrolysis of the phosphoanhydride bonds in ATP. Per mol of ATP hydrolysed, 14.6 kcal of energy are produced. Since cells heavily consume energy, ATP has to be continuously regenerated by mitochondria.

The largest subcellular structure depicted in Fig. 4 is the *nucleus*. The nucleus contains almost the cell's entire DNA. The nucleus is surrounded by a nuclear envelope, composed of two membrane bilayers separated by a narrow space and is continuous with the rough ER. At intervals, the inner and outer nuclear membranes are pinched together and form nuclear pores. Huge protein complexes that allow macromolecules to pass between the cytoplasm and the aqueous phase of the nucleus are associated with these pores [8]. Traffic into the nucleus includes enzymes required for DNA replication, transcription, and RNA processing. Passing out through the nuclear pores are messenger RNA molecules which are translated into protein on ribosomes.

3 Biomolecules

3.1 Molecular Composition of Cellular Membranes

In the following, the basic features of biological membranes of eukaryotic cells will be described. More information on membranes can also be found in chapter B5.

All biological membranes share certain fundamental properties. They are impermeable to most polar or charged solutes, but permeable to nonpolar compounds. Membranes are sheetlike structures. They are 6 to 10 nm thick. Biological membranes consist of a lipid bilayer and a huge amount of either membrane-associated or membrane-integrated proteins. Membrane lipids are relatively small molecules with both, hydrophilic and hydrophobic

features. It is the hydrophobic properties that are relevant for cellular function. The hydrophobic nature originates from a particular constituent of each lipid, i.e. fatty acids. Fatty acids are hydrocarbon chains of various lengths. In most biological membranes fatty acids typically consist of 14 to 24 carbon atoms. Fatty acids not only differ in chain length. They can also be saturated or unsaturated which means that carbon atoms are connected by single or double bonds, respectively.

Three major kinds of lipids can be found in a biological membrane: phospholipids, glycolipids, and cholesterol. *Phospholipids* are abundant in all biological membranes. A phospholipid molecule is constructed from four components: fatty acids, glycerol or sphingosine, a phosphate group, and an alcohol attached to the phosphate. In phosphoglycerides (Fig. 5), the carboxyl groups of two fatty acid chains are esterified to the hydroxyl groups at C-1 and C-2 of glycerol. Phosphoric acid is esterified to the hydroxyl group at C-3 of glycerol. The phosphate can form another ester bond to the hydroxyl group of one of several alcohols. Some common alcohol moieties of phosphoglycerides are choline, ethanolamine, the amino acid serine, and the sugar inositol.

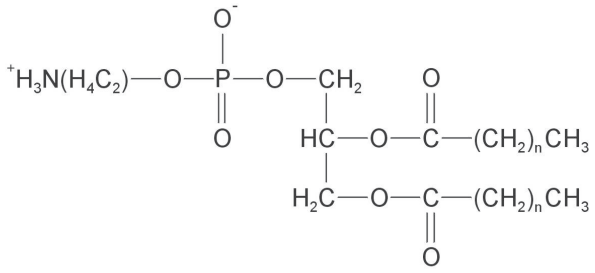


Fig. 5: *Molecular structure of a phosphoglyceride (here: phosphatidylethanolamine) found in cell membranes. Two fatty acid molecules ($CH_3(CH_2)_nCOOH$) are linked by ester bonds to the central alcohol, glycerol (CH_2OH-$CHOH$-CH_2OH). The third hydroxyl group of glycerol is esterified to a phosphate group (PO_4) which also is connected to one of several alcohols, in this case ethanolamine ($^+H_3N(C_2H_4)OH$).*

Glycolipids, as their name implies, are sugar-containing lipids. Glycolipids in animal cells are derived from sphingosine, an amino alcohol that contains a long, unsaturated hydrocarbon chain. One or more sugars are attached to the primary hydroxyl group of the sphingosine backbone. Glycolipids are oriented in a completely asymmetric fashion with the sugar residues always presented on the extracellular side of the membrane.

Cholesterol is another important constituent of biological membranes. It is a steroid. In membranes, cholesterol is oriented parallel to the fatty acid chains of the phospholipids. Cholesterol is found to varying degrees in all animal membranes. A cholesterol content of up to 25% has been determined in some neurons whereas the steroid is essentially absent from intracellular membranes.

What properties enable phospholipids to form membranes? Membrane formation is a consequence of the amphipathic character of the lipid molecules. The polar head groups, formed by the alcohol and phosphate groups favour contact with water (see Fig. 5). Their hydrocarbon chains interact with one another rather than preferring water. These strongly

opposed preferences of the hydrophilic and hydrophobic moieties of membrane lipids can be satisfied by forming a lipid bilayer, composed of two lipid sheets (Fig. 6). The hydrophobic tails (= fatty acid chains) of each sheet interact with one another, forming a hydrophobic interior that acts as a 'natural' permeability barrier for charged or polar molecules. The hydrophilic head groups interact with the aqueous medium on each side of the bilayer. The major driving force for lipid bilayer formation originates from the hydrophobic interactions between the hydrocarbon chains. Van der Waals attractive forces between the hydrocarbon tails favour close packing of these entities. In addition, there are electrostatic and hydrogen-bonding attractions between the polar head groups and water molecules. Thus, lipid bilayers are stabilized by a full array of forces that mediate molecular interaction in biological systems.

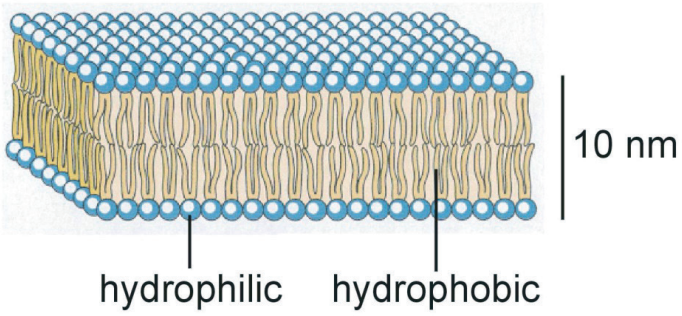


Fig. 6: *Biological membranes consist of a two-dimensional sheet of lipid molecules arranged in a lipid bilayer. The hydrophilic head groups (blue) of the lipids are in contact with water at each surface of the bilayer. The hydrophobic chains of the fatty acids (yellow) are protected from water and form the core of the bilayer. Modified from [2].*

Although the lipid bilayer structure itself is stable, the individual phospholipid molecules have great freedom of motion within the plane of the membrane. The interior of the bilayer is fluid. Individual hydrocarbon chains of fatty acids are in constant motion produced by rotation about the carbon-carbon bonds. The degree of fluidity depends on lipid composition and temperature. At low temperature, lipid motion is restricted and the bilayer exists as a nearly crystalline array. Above a certain temperature (= transition temperature), the crystalline solid changes to a more fluid phase. The transition temperature depends on the lipid composition of the membrane. High amounts of cholesterol and unsaturated fatty acid chains in phospholipids result in lower transition temperatures.

Another type of lipid motion involves the movement of an entire lipid molecule relative to its neighbours. A molecule in one monolayer of the bilayer can diffuse laterally so fast that it 'circumnavigates' a cell within seconds. This rapid lateral diffusion within the plane of the bilayer randomizes the positions of individual molecules in a few seconds. In recent years, however, experimental evidence has been accumulated that some areas of a cell's membrane possess a rather fixed lipid and cholesterol composition. Such areas are referred to as *lipid rafts*. It is assumed that certain membrane proteins preferentially assemble in these areas ('microdomains') to ensure specific and rapid responses to external stimuli.

In addition to lateral movements of lipid molecules one might also expect diffusion from one face of the bilayer sheet to the other. This "flip-flop" transition requires a polar head group to

leave its aqueous environment and move into the hydrophobic interior of the bilayer. This process is energetically unfavourable due to a large, positive change in the free energy. Sometimes, however, a cell necessitates moving lipids from one side of the bilayer to the other. For this purpose, specific proteins in the plasma membrane, called *flippases*, provide a transmembrane path to circumvent the energetically unfavourable flip-flop diffusion.

3.2 The Nucleus

The nucleus contains the majority of the cell's genetic material. Here, multiple linear DNA molecules and specific proteins are tightly associated forming structures called *chromosomes*. Somatic cells that make up most of the human body have two copies of each chromosome. Such cells are 'diploid'. Gametes, i.e. egg and sperm, contain only one copy of each chromosome. These cells are 'haploid'. A human somatic cell has 23 pairs of chromosomes. The DNA, wound up in the chromosomes, contains all the information to encode every protein of the cell. Before a cell starts to divide, the entire chromosomal DNA is duplicated, a process termed *replication*. Thereafter each chromosome consists of two identical *chromatids*. When the cell divides, the two chromatids separate, one moving to each pole of the cell, where they become part of the newly formed nucleus of each daughter cell.

In the nucleus, the DNA is tightly bound to a family of positively charged proteins, the *histones*. Histones and DNA associate in complexes called *nucleosomes*, in which the DNA strand winds around a core of histone molecules. The DNA of a single human chromosome forms about a million nucleosomes. The nucleosomes associate to form very regular and compact supramolecular complexes. The resulting chromatin fibres, about 30 nm in diameter, condense further by forming a series of looped regions and cluster with adjacent looped regions. This tight packing of DNA into nucleosomes results in a remarkable condensation of the DNA molecules. The DNA in the chromosomes of a single somatic human cell would have a combined length of about 2 m if fully extended, whereas the combined length of all 46 chromosomes is about 200 nm (= 2×10^{-7} m). In the following chapter, a short introduction into the molecular properties of the nucleic acids DNA and RNA will be given.

3.2.1 Nucleic Acids

In modern organisms, deoxyribonucleic acids (DNA) and ribonucleic acids (RNA) are the molecular repositories of genetic information. The structure of every protein is a product of information programmed into the nucleotide sequence of a cell's nucleic acids. The ability to store and transmit genetic information from one generation to the next is a fundamental condition of life.

The nucleotide sequence in the cell's DNA specifies the amino acid sequence (see 3.2.2) of every protein in the cell. A segment of a DNA molecule that contains the information required for the synthesis of a functional biological product is referred to as a *gene*. The human *genome* is estimated to contain up to 25,000 different genes. The storage and transmission of biological information are the only known functions of DNA. In comparison to DNA, RNAs have some broader range of functions. Several classes of RNAs are found in cells. Ribosomal RNAs (rRNA) are structural components of ribosomes, the particles that synthesize proteins. Messenger RNAs (mRNA) carry genetic information from a gene to a ribosome, where the encoded protein is synthesized. Transfer RNAs (tRNA) are adapter molecules that translate the information from the messenger RNA into a specific amino acid sequence of a protein.

The nucleic acids of a cell consist of nucleotides (Fig. 7). Nucleotides have three characteristic components: a nitrogen-containing base, a pentose (= sugar), and a phosphate. The bases are derivatives of two parent compounds, purine and pyrimidine. The base of a nucleotide is covalently linked by an N-β-glycosyl bond to the pentose, and the phosphate is esterified to the 5' carbon of the pentose. Both DNA and RNA contain two purine bases, adenine (A) and guanine (G), and two pyrimidines. In both nucleic acids one of the pyrimidines is cytosine (C), whereas the second is thymine (T) in DNA and uracil (U) in RNA. Nucleic acids have two kinds of pentoses. The nucleotides of DNA contain 2'-deoxy-D-ribose, and the nucleotides of RNA contain D-ribose.

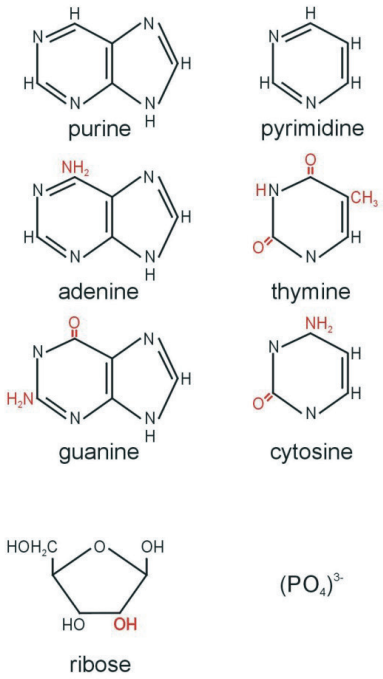


Fig. 7: *Chemical components of nucleotides. Nucleotides consist of three components: a nitrogen-containing base, a pentose, and a phosphate. The bases are derivatives of either purine (adenine or guanine) or pyrimidine (thymine or cytosine). In DNA, the hydroxyl group on the pentose ring (depicted in red) is replaced by hydrogen.*

The successive nucleotides of both DNA and RNA are covalently linked through phosphate-group bridges, in which the 5'-hydroxyl group of one nucleotide is joined to the 3'-hydroxyl group of the next nucleotide by a phosphodiester linkage (Fig. 8). Thus the covalent backbones of nucleic acids consist of alternating phosphate and pentose residues, with the bases joined as side groups to the backbone at regular intervals. The backbones of both DNA and RNA are hydrophilic. The hydroxyl groups of the sugar residues form hydrogen bonds with water. The phosphate groups are completely ionized and negatively charged at pH 7. Thus, DNA is considered as a 'poly-anion'. In a cell, the negative charges are compensated by ionic interactions with positive charges on proteins, e.g. histones, metal ions, or polyamines.

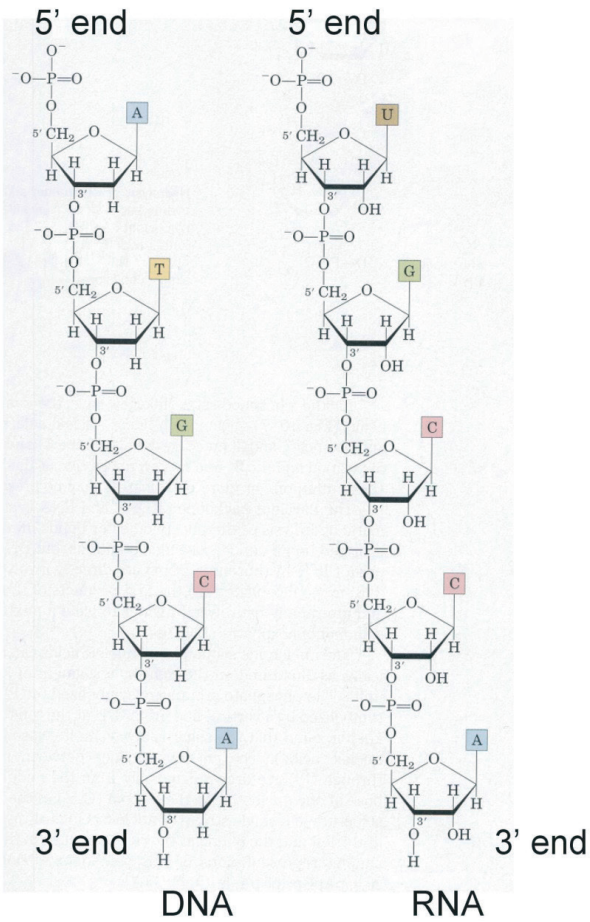


Fig. 8: *Covalent bonds in the backbone of DNA and RNA. The phosphodiester bonds link successive nucleotide units. The backbone of alternating pentose and phosphate groups in nucleic acids is highly polar. Sketch modified from [2].*

The covalent structure of nucleic acids accounts for their ability to carry information in form of the linear sequence of bases along a nucleic acid chain. Another important feature is the ability of the bases to form specific base pairs (A : T; G : C) which gives rise to a helical structure consisting of two complementary strands of nucleic acids. The existence of specific base-pairing interactions was discovered when the three-dimensional structure of DNA was solved by X-ray diffraction [9]. The diffraction patterns indicated that DNA was formed of two chains that wound in a regular helical structure. The two helical DNA chains are coiled around a common axis. The sugar-phosphate backbones are on the outside of the helix, and the purine and pyrimidine bases are facing the inside of the helix. The bases are nearly perpendicular to the helix axis. The helical structure repeats every 3.4 nm. There are 10 base pairs per helix turn. The diameter of the helix is 2.0 nm. Altogether these are the essential features of the Watson-Crick model of DNA.

The flow of genetic information in a cell is from DNA to RNA to protein. All cellular RNAs are synthesized by specific enzymes (RNA polymerases) from DNA as a template. This process is called *transcription*. In contrast to DNA molecules, RNA molecules are single stranded. RNAs, however, may also contain double-stranded regions that arise from the folding of the single chain into hairpin loop-like structures.

3.2.2 Synthesis of proteins

Proteins are the 'working horses' of a cell. Proteins are constructed from 20 different *amino acids*. All amino acids share a basic design. A central carbon atom (C_α) is bonded to an amino group, to a carboxyl group, to a hydrogen atom, and to one variable group (R), called side chain (Fig. 9). It is the side chain that determines an amino acids physico-chemical property. Amino acids are the letters of the alphabet in which proteins are "written".

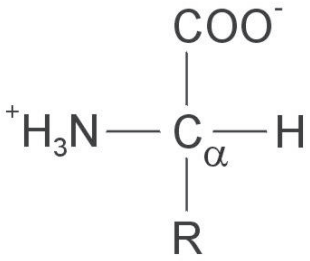


Fig. 9: *Structural formula of amino acids.*

Transcription of a gene by RNA polymerase II results in the synthesis of messenger RNA (mRNA). The mRNA is the template for protein biosynthesis. How does an mRNA molecule direct amino acids to become joined in the correct sequence to form a protein? The process of protein biosynthesis requires another class of RNA molecules, i.e. transfer RNA (tRNA). All tRNAs contain an amino acid-attachment site and a template (= mRNA) recognition site. The carboxyl group of the amino acid is esterified to the 3'- or 2'-hydroxyl group of the ribose at the end of the tRNA chain (= amino-acyl-tRNA). This reaction is catalyzed by specific enzymes (= synthetases) and there exist specific synthetases for attaching each of the 20 different amino acids to their respective tRNA(s).

The correct arrangement of 20 different amino acids in a protein cannot be specified by four nucleotides in a one-to-one manner. Consequently, a group of nucleotides is required to encode each amino acid. The code employed must be capable of specifying at least 20 'words', i.e. different amino acids. If two nucleotides were used to code for one amino acid, then 4^2 different 'words' could be formed; a number still below the default value. If three nucleotides are used to encode one amino acid, then 4^3 code 'words' can be formed. This condition not only fulfils but also exceeds the default value. In fact, some amino acids are encoded by more than one triplet codon. This phenomenon is referred to as '*the genetic code is degenerate'*. An advantage of the degeneracy is that the system becomes less vulnerable to mutations in the nucleic acids. Even the DNA base composition may vary without altering the amino acid sequence of the encoded protein. The template-recognition site on tRNA is a sequence of three nucleotides called *anticodon*. The anticodon on a tRNA molecule recognizes a complementary sequence on the mRNA by base pairing.

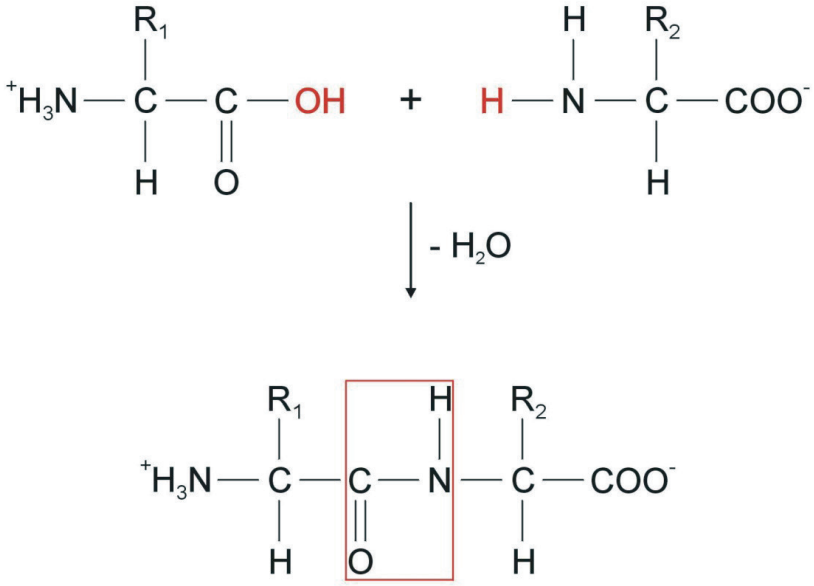


Fig. 10: *Schematic drawing of the formation of a peptide bond. Proteins are formed by linking the α-carboxyl group of one amino acid (R1) to the α-amino group of another amino acid (R2) with a peptide bond. The formation of a dipeptide from two amino acids is accompanied by the loss of a water molecule.*

Ribosomes are the molecular machines coordinating the interplay of mRNA, tRNAs, and proteins which leads to protein synthesis. A eukaryotic ribosome consists of a large and a small subunit. Both subunits come together forming the functional ribosome with a relative molecular mass of 4,200 kDalton. As a whole, the ribosome is a complex composed of individual RNA molecules (rRNA) and more than 50 proteins [10].

The average rate of protein synthesis is about five amino acids per second. At the end of biosynthesis, the protein is a linear stretch of covalently linked amino acids (see Fig. 10). To become a functional protein within the cell it has to adopt a unique three-dimensional structure. A protein adopts a stable, folded conformation mainly through noncovalent ionic, hydrogen, van der Waals, and hydrophobic interactions. Its stability is further enhanced by disulfide bonds formed between amino acid residues harbouring reactive sulfhydryl (SH) groups. Detailed information on protein folding can be found in chapters E2 and E4.

3.2.3 Cell regulation by epigenetic factors

Eukaryotic cells are equipped with and make use of rather similar molecular components and reactions to meet daily demands. Nevertheless, cells in the human body are phenotypically and functionally quite distinct and specialized although they share the same genetic background. Currently, an intensive field of investigation in cell biology aims to unravel processes that lead to cellular specialization at the molecular level. Whereas the general structure of chromatin is similar in all eukaryotic cells, chromatin has to be considered as a dynamic structural and functional scaffold governing processes such as transcription, DNA

replication, and cellular differentiation. These processes are influenced by a multitude of genetic and *epigenetic* factors. DNA methylation and post-translational modification of histones are prominent epigenetic factors. In eukaryotic organisms these epigenetic events have evolved to provide a precise and stable control of gene expression and genomic regulation even through multiple cycles of cell division. Deregulation of epigenetic factors and events may lead to diseases whereas the loss of epigenetic heritance mechanisms may result in cellular aging and cancer.

Histones that participate in building the nucleosomes are subject to multiple post-translational, often transient modifications that include acetylation, methylation, phosphorylation and sumoylation. The particular combination of histone post-translational modifications constitutes a so called 'histone code' that influences chromatin structure and function by altering nucleosome-nucleosome interactions as well as interactions with various proteins that participate in or regulate processes from gene expression to genetic recombination.

Histone acetylation and deacetylation are reversible reactions mediated by specific enzymes. In the deacetylated form, basic amino acids in histones are positively charged and interact with DNA's negatively charged phosphate groups or negatively charged patches on neighbouring nucleosomes. These interactions promote chromatin condensation. In contrast, acetylation of those residues neutralizes these positive charges and promotes a less condensed and more accessible *chromatin* conformation. Thus, histone acetylation has been linked to actively transcribed genomic regions. Histone acetylases and histone deacetylases, the enzymes adding or hydrolyzing the acetyl-groups, have been traditionally linked to transcriptional activation and repression, respectively. Furthermore, transient modifications have been implicated in many biological processes, such as cell survival and differentiation, cell cycle progression, tumorigenesis, cardiac function and remodelling in health and disease. At present, histone acetylases and deacetylases are among the most promising targets in cancer therapy, and the search for specific modulators of their activities has attracted much attention.

In addition to acetylation, methylation of basic amino acid residues in histones has been examined extensively in the recent past. Similar to the transfer of acetyl groups, methylation of amino acid residues can contribute to active or repressive chromatin functions. Notably, histone methylation has been implicated in the maintenance of embryonic stem cells in the undifferentiated state and thus to play a key role in differentiation. From a chemical point of view, histone methylation is more stable than acetylation or phosphorylation and has been considered as a rather static modification. Recently, however, specific enzymes have been identified that catalyze the removal of methyl groups from histones. Therefore, methylation of histones may turn out as a transient modification that is as reversible as acetylation.

Besides on proteins, methylation can also occur on the DNA itself. In mammalian cells, specific enzymes, cytosine-5'methyltransfereases, add a methyl group to the C-5 atom of cytosine. The methylation of DNA is an important epigenetic regulator of gene expression. In the mammalian genome, about 70% of the CpG dinucleotides are methylated. Many of the remaining non-methylated CpGs are located in 'CpG islands' typically found in functional *promoter* regions. DNA methylation has long been considered as a marker of gene repression which plays important roles in long-term silencing of repetitive elements or X-chromosome inactivation.

Recent studies, however, show that transcriptional activation is associated with cycles of DNA methylation and that methyltransferases participate in both the addition and removal of methyl groups. Whereas methylation of histones keeps embryonic stem cells in an undifferentiated state, differentiation of these cells is associated with changes in DNA methylation. Hence, the DNA methylation pattern in CpG islands is among the factors that control the differentiation of the various cell types in the body. Furthermore, tumorigenesis is frequently associated with hypermethylation of CpG islands in promoters of tumor suppressor genes. Thus, the DNA methylation states and the diverse combinations of the different histone post-translational modifications lead to a variety of functional consequences for an organism. It is the epigenetic factors that can have a profound impact on gene expression which in turn may control either normal development or progression of disease.

4 Summary

This chapter aimed to introduce the reader to some of the morphological and molecular properties of eukaryotic cells. Subcellular structures like the nucleus and cellular membranes as well as important biomolecules have been addressed. Pivotal to all cellular functions is the presence and activity of specific proteins. Proteins are synthesized from rather simple monomeric units, the amino acids. Proteins acquire their functional activity after folding in a characteristic three-dimensional structure. To a certain extent, the structural features of a protein are already determined by the genetic information stored in the DNA. The DNA is wound up as a nucleic acid/protein complex in the chromosomes. Transcription of a gene produces molecular messengers required to "translate" the genetic information into a protein sequence. As a general rule, the flow of genetic information is from DNA → RNA → protein, but proteins in turn synthesize both DNA and RNA. All cellular processes need a continuous supply of energy. In biological systems, most of the energy is gained from hydrolysis of ATP. Mitochondria ensure the constant supply of ATP as it is these cellular organelles where aerobic degradation of sugar or fatty acids to CO_2 and H_2O is coupled to ATP production. Though the biochemical and molecular mechanisms described in this section are known for reasonable time, biochemists and cell biologists are facing novel challenges. The advent of completely sequenced genomes led to the prediction of protein sequences with hitherto unknown function. Analysing these proteins and unravelling their properties at the cellular level is a tremendous challenge in forthcoming years. In addition, an increasingly diverse picture is emerging of a range of epigenetic mechanisms participating in cellular regulation. In this field of cell biology, the development of techniques that employ genetic and chemical means to achieve cellular reprogramming and/or epigenetic modification in stem and somatic cells may lead to novel and promising approaches to treat life-threatening conditions like cancer or different forms of neurodegeneration, e.g. Alzheimer's or Parkinsons's disease.

References

[1] Berg, J.M., Tymoczko, J.L., and Stryer, L. "*Biochemistry*" 5th Ed. (W.H. Freeman & Co., New York, 2001). See especially chapters 5, 12, 14, and 29.

[2] Nelson, D.L. and Cox, M.M. "*Lehninger Principles of Biochemistry*" 5th Ed. (Worth Pub., New York, 2000). See especially chapters 2, 10, 11, and 19.

[3] Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., and Bretscher, A. "*Molecular Cell Biology*" 6th Ed. (W.H. Freeman & Co., New York, 2007).

[4] Alberts, B., Johnson, A., Walter, P., Lewis, J., Raff, M., Roberts, K., and Orme, N. "*Molecular Biology of the Cell*" 5th Ed. (Garland Science Pub. Inc., New York, 2007).

[5] Lottspeich, F. and Engels, J.W. "*Bioanalytik*" 3rd Ed. (Springer Spektrum Verlag, Heidelberg, 2012).

[6] Gilbert, S.F. and Singer, S.R. "*Developmental Biology*" 8th Ed. (Palgrave Macmillan Ltd, Barsingstoke, 2006).

[7] Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., and Hudspeth, A.J. "*Principles of Neural Science*" 5th Ed. (Mcgraw-Hill Pub. Comp., New York, 2012).

[8] Strambio-De-Castillia, C., Niepel, M. and Rout, M.P. (2010) The nuclear pore complex: bridging nuclear transport and gene regulation. Nat. Rev. Mol. Cell. Biol. 11, 490-501.

[9] Watson, J.D. "*The Double Helix: A Personal Account to the Discovery of the Structure of DNA*" (Norton & Co. Inc., New York, 1981).

[10] Moore, P.B. (1998) The three-dimensional structure of the ribosome and its components. Annu. Rev. Biophys. Biomol. Struct. 27, 35-58.

A 3 Biohybrids

T. Weidner
Max-Planck-Institut for Polymer Research
Mainz

Contents

1 Introduction 2

2 Applications 2

3 Macroscopic Observations of Biosilification 5

4 Understanding Protein Control of Biomineralization 6

5 Closing Remark 8

References 9

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

The precise control of solid inorganic structures remains a challenge for material scientists and engineers. Nano- and micro-fabrication of inorganic materials generally requires elaborate catalysts and extremes in temperature, pressure and pH, along with the well-known risks for health and the environment. In contrast, natural organisms have developed the ability to shape complex inorganic materials, mineral biohybrids, at moderate temperatures, and near neutral pH, often in aqueous environments.[1] Proteins are particularly important for biological organisms to control the shape, mechanical properties and morphology of mineralized tissues. Proteins are essentially Nature's engineers of hard materials – they control the growth of inorganic composites containing for example hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$) the major component of bone and teeth, calcium carbonates ($CaCO_3$) in sea urchins, and silica ($SiO_2 \cdot {}_nH_2O$ in diatoms, marine sponges and some plants (Figure 1). [2-7]

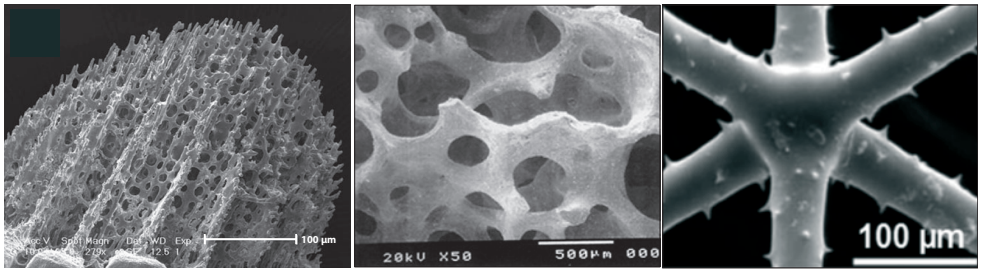


Fig. 1: *Mineralized tissue generated by protein control. Left: Spine of a sea urchins consisting of calcium carbonates. Middle: bone hydroxyapatide. Right: silica spicule of the sponge M. chuni. Figure based on Refs. 2, 4-7.*

2 Applications

Mimicking hard tissue formation has led to numerous potential applications in bioengineering. One strategy has been to coat biomaterials such as hip replacements, with hydroxyapatite composites. Hydroxyapatite is the main constituent of bone and teeth and is therefore 'naturally' biocompatible.[6] Hydroxyapatite coatings can promote the biocompatibility and osseointegration of bone and teeth replacements and other implants.

When it became clear that adhesion proteins of mussels, polydopamines, can promote the formation of hydroxyapatite films, many research groups started experimenting with polydopamine for the production of biocompatible surfaces and particles.[8] Figure 2A shows an example for a calcium phosphate particle grown on a polydopamine-coated surface.[9]

Our understanding of the interaction of surface bound peptides with molecules in solution is still limited. Currently there is a growing interest in understanding the molecular level detail of these processes and finding 'design rules' to improve the mineralization efficiency and control the phase and morphology of minerals grown.

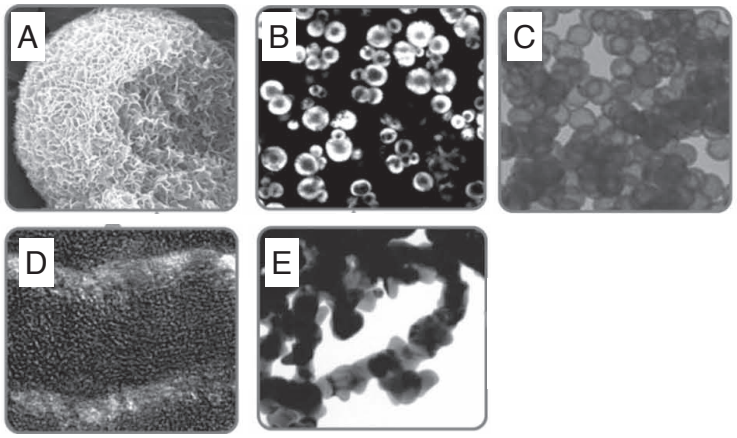


Fig. 2: *Application of protein-based mineral biohybrids. (A) Bone replacement material inspired by mussel tissue. Image: 5x5 μm².[9] (B). Hollow $CaCO_3$ microspheres from CO_2 conversion. Image: 60x60 μm².[12] (C) Anti-cancer doxorubicin-loaded silica particles. Image: 100x100 μm².[13] (D) TiO_2-based composite material inspired by nacre. Image: 250x250 nm².[14] (E) Anodes for Li-ion batteries fabricated by genetically engineered viruses. Image: 200x200 nm².[15]*

Specific biomineralization pathways in mollusks and corals lead to the storage of vast amounts of CO_2 in solid inorganic materials. Current efforts to store climate changing carbon dioxides make biogenic CO_2 storage a promising field of study.[10] In addition to providing convenient pathways towards CO_2 storage, the CO_2 induced mineralization in corals can also yield high fidelity microstructures in the lab.[11] When Kim et al. injected CO_2 into a mixture of $CaCl_2$ and dopamine (an adhesive protein in mussels) they observed the formation of hollow $CaCO_3$ microspheres (Figure 2B).[12] In addition, the spheres could be reacted to form hydroxyapatite and be deposited onto bone bioactive scaffolds for novel implant surfaces.

Porous hydroxyapatite and $CaCO_3$ have been used to store and release anti-inflammatory drugs, which are especially useful for treating bone infections. In addition, capsules of calcium carbonates have been used as vessels for the delivery of anti-cancer drugs.[13] The structure of calcium carbonates is pH-dependent. Since cancer cells typically maintain a lower pH compared to healthy cells, the pH-dependence makes is possible to selectively deliver drug cargoes such as the anti cancer drug doxorubicin (DOX) to cancer cells (Figure 2C). The idea is that the carrier particles will disintegrate or change structure and release the cargo once it is in an acidic cell environment.

In Nature, many composite materials comprised of hard and soft tissue exhibit very high toughness. These natural materials can be flexible, transparent and still rival steel in their mechanical properties. The high toughness is often generated by exquisite control of a layered material of alternating soft and hard tissues, which are effectively ‘glued’ together so that stress is transferred between the layers. While in Nature high fidelity layers are typically fabricated and controlled by proteins, researchers have mimicked the architectures with polymers such as, for example, poly acrylic acid. The polymers’ COOH functionalities approximate the charge density of glutamine-rich proteins and polypeptides. Figure 2D shows such a hybrid material – titanium dioxide layers sandwiched between oppositely charged

organic films comprised of poly(styrenesulfonate), polyethyleneimine and poly(allylamine hydrochloride) assembled by a layer-by-layer method.[14] The organic/inorganic thickness ratio is a critical parameter for the toughness of the material. In this example the best mechanical performance was observed for a ratio of 1/10, interestingly a value that is also typical for nacre. The extremely thin organic layers required for high material performance are a challenge for material synthesis and self-assembled layers based on proteins will likely play an important role in future concepts.

In batteries the electrode material is typically made of inorganic oxides and phosphates such as lithium oxide for Li-ion batteries. Recently, effective biomimetic anode materials have been produced by templating with M13 viruses (Fig.2E).[15] The viruses were genetically modified to carry two additional kinds of peptides, one with high affinities for carbon nanotubes and another for $FePO_4$. The resulting composite material was stable and performed comparably to state of the art materials. The benefit of biomimetic synthesis is the improved energy efficiency and lower biological impact. On the other side of the complexity scale, besides viruses, researchers have also used extremely small and simple homopolypeptides such as diphenylalanine as building blocks for nanostructured electronic materials.

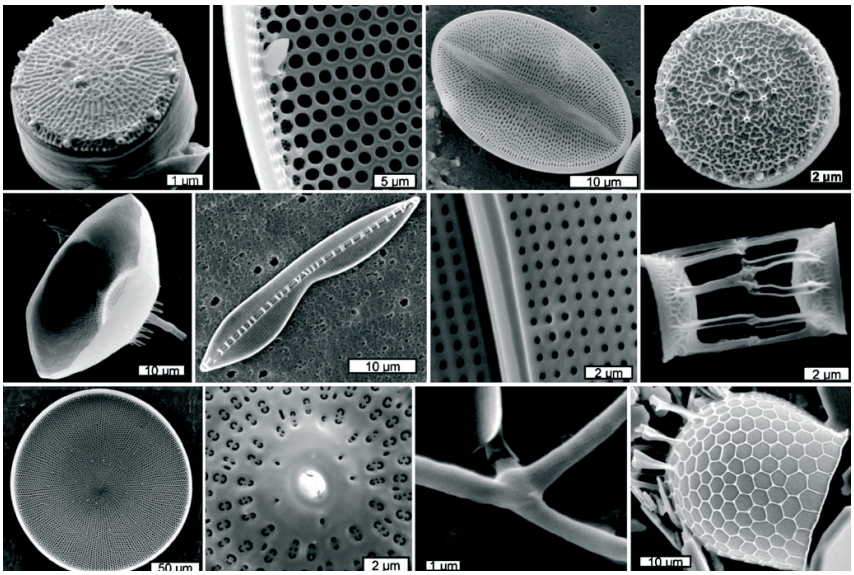


Fig. 3: *Acid-cleaned biosilica cell walls illustrating the morphological diversity of silica frustules in different diatom species. Figure from ref 3.*

The most abundant biomineral is silica – $SiO_{2-n}H_2O$.[3] The large interest in biogenic silica formation, i.e. biosilicification, is driven by the numerous potential applications of silica. The fabrication of devices using biogenic silification will require precise structural control. Applications include silica-based optical fibers and drug delivery systems derived from mesoporous silica.

A large variety of species including marine sponges, radiolarians, and diatoms are capable of producing complex silica structures, which derive from orthosilicic acid – $Si(OH)_4$. Diatoms are the dominant siliceous species, they produce intricate silica-based cell walls (frustules, see

Fig.3). Besides producing ca. 20% of the atmospheric oxygen they fabricate the largest portion of the earth's biogenic silica, which in total is estimated at 240×10^{12} moles of silica per year.

As with other biogenic minerals, proteins play a crucial role in the regulation and control of silica formation and construction in diatoms. The biogenic silica of diatom cell walls is not entirely inorganic, it also contains a significant fraction of proteins. An important discovery was the observation that the dominant protein fraction in cell walls of the diatom species *Cylindrotheca fusiformis* consists of small cationic proteins with high binding affinity for silica.

The proteins, called silaffins, consist of a series of amino acid repeat sequences called R1-R7. These repeat domain peptides contain large numbers of serines and the basic amino acids lysine and arginine. *In vivo* the peptides also carry a range of post-translational modifications: the serines are predominantly phosphorylated while the lysines are typically alkylated. The current picture of silica cell wall biogenesis takes place in the so-called silica deposition vesicles (SDVs). The process begins with the formation of silica nanospheres, which contain high levels of silaffins. This silica sphere formation has also been observed *in vitro*, outside of SDVs. This discovery has been a starting point of biomimetic silica production in the lab because R peptides with the ability to induce silica formation from silicic acid solutions in vitro are often as small as 19 amino acids and can therefore conveniently be synthesized *de novo* using robotic solid phase synthesis for extended experimentation.
The best studied of the repeat domains is the nineteen amino acid R5 peptide (NH2SSKKSGSYSGSKGSKRRILCOOH).[16] R5 generates silica spheres with an average radius of 500 nm when exposed to silicic acid solutions.

Especially the C- terminal RRIL sequence of has been shown to be critical for silica formation, and mutants with RRIL deleted or altered, do not precipitate silica.[17] In addition, R5 can also precipitate silica with its serine and lysine sites unmodified (i.e. without phosphorylation and alkylation) but the process requires higher pH than for native silaffins or R5 having modified side chains.

3 Macroscopic Observations of Biosilification

The macroscopic aspects of biosilica morphology and the role of peptide composition and external influences such as shear and flow have been studied extensively.[18] It is now clear that the morphology of silica precipitated by the R5 peptide can be altered by the application of shear stress and other externally applied forces. Additional variations of the amino acid sequence lead to a variety of differently shaped particles, including spheres, rods, tubes, wires and sheets. As mentioned above, even peptides as small as the tetrapeptide I3K can induce the precipitation of silica nanotubes.[19] Interestingly the closely related L3K peptide –there is a small change of the positions of methyl units between leucine (L) and isoleucine (I) – precipitates nanospheres instead of tubular structures.

A variety of cationic long chain polyamines also induce silica formation. The identification of new proteins capable of precipitating silica structures is an ongoing effort. New proteins and biomolecules are either identified by proteomic analysis of mineralizing organisms and by trial and error-type screening of libraries. Currently there in increased effort to understand the

mechanisms underlying protein regulated mineralization for a more rational design approach. A very impressive example is the control of calcite crystal shapes by synthetic peptides design using computer simulations as demonstrated by Grey and coworkers.[20] But rational designs are the still the exception, not the rule. Part of the challenge is our limited understanding of how proteins interact with hard material surfaces. To overcome this problem, several groups are developing spectroscopic and computational methods to study interfacial protein structure.

4 Understanding Protein Control of Biomineralization

Despite importance of the interactions between proteins and minerals, and protein control of silica precipitation in particular, due to the difficulties in studying protein structure and function at inorganic and solid polymer surfaces, or within mineral-protein composites, there is still remarkably little known of the molecular mechanisms governing hard tissue generation. Membrane proteins, for example, are much better understood than proteins in contact with hard tissue. Of the roughly 100,000 protein structures accumulated in the Protein Data Base (PDB) by X-ray crystallography, NMR and other techniques, there is still no experimentally determined structure for a protein interacting with its native mineral surface. The Drobny group has developed solid-state NMR (ssNMR) methods to determine the local structure and dynamics of small salivary proteins bound to hydroxyapatite crystals.[21]

The Gray group has used computational methods in combination with ssNMR-derived distance and torsion angle constraints to obtain a structural model for the mineral protein statherin bound to HAP crystals. The human salivary protein statherin can track down and bind specific faces of hydroxyapatite and is thereby involved in controlling the growth rate of different hydroxyapatite crystal surfaces. This control of hydroxyapatite shape and morphology is still mystery – understanding the underlying concepts would allow material scientists to control mineral synthesis with high fidelity.

A first experimentally validated atomic resolution structure of statherin on hydroxyapatite, determined with the structure prediction algorithm RosettaSurface in combination with experimental data is shown in Fig.4.[22] This first-of-its-kind structure of a protein at a biomineral interface shows how statherin is folded and the amino acid side chains involved in binding the hydroxyapatite surface atoms. ssNMR experiments use proteins bound to micro- or nanoparticles, which can be very different from a flat surface.

Weidner and coworkers have developed high-resolution spectroscopic methods to probe the structure of small peptide on surface directly.[23, 24] Using surface sensitive sum frequency generation (SFG) spectroscopy in combination with near-edge X-ray absorption fine structure (NEXAFS) spectroscopy they determined the orientations of specific statherin sites with respect to the surface.[25] Together, this first high-resolution structural information is a step towards understanding, for example, the structural basis for the facet specificity of statherin.

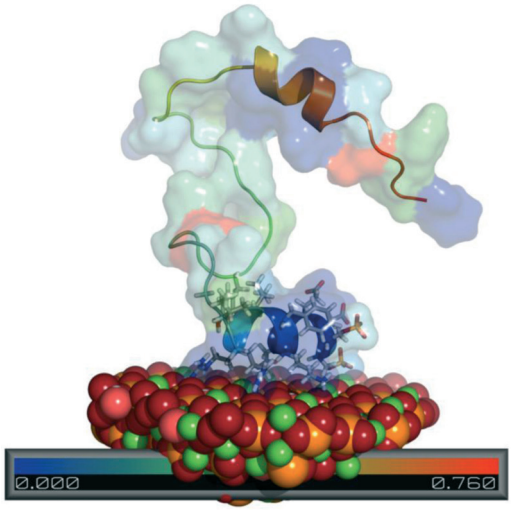


Fig. 4: *Structure representative the 100 structures with lowest energy for statherin on hydroxyapatide. The color bar shows deviations in angstroms in the opaque profile surrounding the protein.*[22]

In order to understand the concepts used by diatoms to efficiently fabricate nanostructured silica it became clear early on that simplified model systems for silification will be necessary for studies at the molecular level – the silification machinery in diatoms and related species is extremely complex.[3] Drobny and co-workers found that artificial peptides consisting of lysine and leucine (LK peptides) can mimic the diatoms capability of forming various biosilica nanostructures.[26] These extremely simple peptides are designed to adopt helical or beta-sheet structures are therefore ideal model systems to study the effect of folding on silica morphology (Fig.5).

The SEM images show how LK peptides folded differently at the water-silica interface can generate different silica morphologies – ranging from spheres to rods and wire-type structures. The folding and aggregation on silica within silica particles shown in the figure was determined using sum frequency generation vibrational spectroscopy in combination with ssNMR and molecular dynamics simulations.[27]

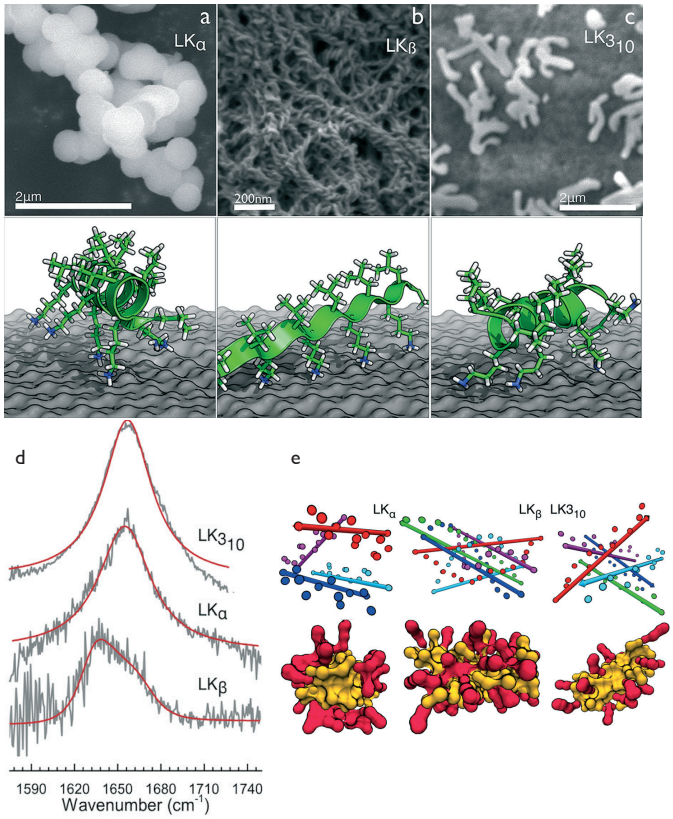


Fig. 5: *a-c: Precipitated silica structure depends on LK peptide folding in contact with a silica surface. d: SFG amide I spectra used to determine the peptide folding on silica. e: MD simulation snapshots of LK peptide interaction with silica in solution.*[27]

5 Closing Remark

The advent of rapid, mass spectrometry-based proteomic methods will allow the identification of a host of previously unknown proteins involved in biogenic mineralization.

Newly developed spectroscopies will allow researchers to study the mode of action by which proteins interact with and control mineral growth. Taking clues from Nature, material scientists and engineers continue to learn important lessons towards the goal of rational, biomimetic synthesis of functional and environmentally friendly organic-inorganic biohybrid materials.

References

(1) Kim, S.; Park, C. B., Bio-Inspired Synthesis of Minerals for Energy, Environment, and Medicinal Applications. *Adv. Funct. Mater.* **2013,** *23*, 10-25.

(2) LeGeros, R. Z., Calcium Phosphate-Based Osteoinductive Materials. *Chem. Rev.* **2008,** *108*, 4742-4753.

(3) Hildebrand, M., Diatoms, Biomineralization Processes, and Genomics. *Chem. Rev.* **2008,** *108*, 4855-4874.

(4) Weiner, S. A., Lia, Design strategies in mineralized biological materials. *Journal of Materials Chemistry* **1997,** *7*, 689-702.

(5) Ma, Y.; Aichmayer, B.; Paris, O.; Fratzl, P.; Meibom, A.; Metzler, R. A.; Politi, Y.; Addadi, L.; Gilbert, P. U. P. A.; Weiner, S., The grinding tip of the sea urchin tooth: exquisite control over calcite crsytal orientation and Mg distribution. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 6048-6053.

(6) Mahamid, J.; Addadi, L.; Weiner, S., Crystallization Pathways in Bone. *Cells Tissues Organs* **2011,** *194*, 92-97.

(7) Weiner, S.; Addadi, L., Crystallization Pathways in Biomineralization. *Annual Review of Materials Research, Vol 41* **2011,** *41*, 21-40.

(8) Kim, S.; Park, C. B., Dopamine-induced mineralization of calcium carbonate vaterite microspheres. *Langmuir* **2010,** *26*, 14730-6.

(9) Kim, S.; Park, C. B., Mussel-inspired transformation of CaCO3 to bone minerals. *Biomaterials* **2010,** *31*, 6628-34.

(10) Feely, R. A.; Sabine, C. L.; Lee, K.; Berelson, W.; Kleypas, J.; Fabry, V. J.; Millero, F. J., Impact of anthropogenic CO2 on the CaCO3 system in the oceans. *Science* **2004,** *305*, 362-6.

(11) Lee, H.; Scherer, N. F.; Messersmith, P. B., Single-molecule mechanics of mussel adhesion. *P. Natl. Acad. Sci. USA* **2006,** *103*, 12999-3003.

(12) Kim, S.; Ko, J. W.; Park, C. B., Bio-inspired mineralization of CO2 gas to hollow CaCO3 microspheres and bone hydroxyapatite/polymer composites. *Journal of Materials Chemistry* **2011,** *21*, 11070-11073.

(13) Wei, W.; Ma, G. H.; Hu, G.; Yu, D.; McLeish, T.; Su, Z. G.; Shen, Z. Y., Preparation of hierarchical hollow CaCO3 particles and the application as anticancer drug carrier. *J. Amer. Chem. Soc.* **2008,** *130*, 15808-10.

(14) Burghard, Z.; Zini, L.; Srot, V.; Bellina, P.; Aken, P. A. v.; Bill, J., Toughening through Nature-Adapted Nanoscale Design. *Nano Letters* **2009,** *9*, 4103-4108.

(15) Lee, Y. J.; Yi, H.; Kim, W. J.; Kang, K.; Yun, D. S.; Strano, M. S.; Ceder, G.; Belcher, A. M., Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes. *Science* **2009,** *324*, 1051-5.

(16) Kröger, N.; Lorenz, S.; Brunner, E.; Sumper, M., Self-Assembly of Highly Phosphorylated Silaffins and Their Function in Biosilica Morphogenesis. *Science* **2002,** *298*, 584-586.

(17) Knecht, M. R.; Wright, D. W., Functional analysis of the biomimetic silica precipitating activity of the R5 peptide from Cylindrotheca fusiformis. *Chem. Commun.* **2003**, 3038-9.

(18) Naik, R. R.; Whitlock, P. W.; Rodriguez, F.; Brott, L. L.; Glawe, D. D.; Clarson, S. J.; Stone, M. O., Controlled formation of biosilica structures in vitro. *Chem. Commun.* **2003**, 238-9.

(19) Xu, H.; Wang, Y.; Ge, X.; Han, S.; Wang, S.; Zhou, P.; Shan, H.; Zhao, X.; Lu, J. R., Twisted Nanotubes Formed from Ultrashort Amphiphilic Peptide I3K and Their

Templating for the Fabrication of Silica Nanotubes. *Chemistry of Materials* **2010,** *22*, 5165-5173.

(20) Masica, D. L.; Schrier, S. B.; Specht, E. A.; Gray, J. J., De Novo Design of Peptide-Calcite Biomineralization Systems. *J. Amer. Chem. Soc.* **2010,** *132*, 12252-12262.

(21) Long, J. R.; Shaw, W. J.; Stayton, P. S.; Drobny, G. P., Structure and dynamics of hydrated statherin on hydroxyapatite as determined by solid-state NMR. *Biochemistry* **2001,** *40*, 15451-15455.

(22) Masica, D. L.; Ash, J. T.; Ndao, M.; Drobny, G. P.; Gray, J. J., Toward a structure determination method for biomineral-associated protein using combined solid- state NMR and computational structure prediction. *Structure* **2010,** *18*, 1678-87.

(23) Weidner, T.; Breen, N. F.; Li, K.; Drobny, G. P.; Castner, D. G., Sum frequency generation and solid-state NMR study of the structure, orientation, and dynamics of polystyrene-adsorbed peptides. *P. Natl. Acad. Sci. USA* **2010,** *107*, 13288-13293.

(24) Weidner, T.; Dubey, M.; Breen, N. F.; Ash, J.; Baio, J. E.; Jaye, C.; Fischer, D. A.; Drobny, G. P.; Castner, D. G., Direct Observation of Phenylalanine Orientations in Statherin Bound to Hydroxyapatite Surfaces. *J. Amer. Chem. Soc.* **2012,** *134*, 8750-8753.

(25) Weidner, T.; Castner, D. G., SFG analysis of surface bound proteins: a route towards structure determination. *Physical chemistry chemical physics : PCCP* **2013,** *15*, 12516-24.

(26) Zane, A. C.; Michelet, C.; Roehrich, A.; Emani, P. S.; Drobny, G. P., Silica Morphogenesis by Lysine-Leucine Peptides with Hydrophobic Periodicity. *Langmuir* **2014,** *30*, 7152-7161.

(27) Baio, J. E.; Zane, A.; Jaeger, V.; Roehrich, A. M.; Lutz, H.; Pfaendtner, J.; Drobny, G. P.; Weidner, T., Diatom mimics: directing the formation of biosilica nanoparticles by controlled folding of lysine-leucine peptides. *J Am Chem Soc* **2014,** *136*, 15134-7.

A 4 Functional Polymers at Flat Interfaces

M. Möller
DWI Leibniz Institute for Interactive Materials
Forckenbeckstr. 50, 52062 Aachen

Contents

1 Introduction 2

2 Polymers, Liquid Crystals, and SAMs 3

3 Adsorption/Desorption of Brush-like Macromolecules 7

4 Microgel Adsorption, Molecular-like or Colloid-like 8

5 References 14

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

Surface induced ordering and structure formation is of great practical importance in many technical applications. Important examples are the orientation of liquid crystals in displays, controlling wetting of coating formulations with the aid a surfactant, regular nanostructure and lateral pattern formation in thin blockcopolymer films, and the formation of highly ordered monolayers of colloids. In the present paper we will address specific ordering effects and pattern formation of macromolecules at an interface with an emphasis on the tools that are offered by tailored variations of the polymer structure including their conformational flexibility, selfassembly and the introduction of functional groups that undergo specific and directed interaction.

It is well known that an interface imposes local ordering to the molecules in a liquid that is in contact with it merely by the topography of the interaction forces [1], similar like near range ordering occurs in a liquid [2.3]. For low molecular weight compounds, typically the surface structure is confined to distances no more than one or two molecular lengths from the interface. In the case of a low molecular weight liquid crystalline compound, the interface can induce structure that penetrates hundreds of molecular lengths into the bulk [4]. Deposition of a layer of hard colloidal particles from an evaporating solution is in addition to the particle-surface and particle-particle interaction also controlled by meniscus forces in the thinning liquid film and has been shown to yield highly ordered 2D colloidal crystals [5]. Monolayer formation with surfactant type molecules is typically controlled by the specific interaction of a head or tail group with the second phase forming the interface and the lateral interaction within the monolayer, examples are Langmuir-Blodgett films [6] and self-assembled monolayers [7].

For the case of polymer molecules, i.e. flexible chain molecules, either linear or branched, and in the case of microgels constituting an open polymer structure with internal crosslinks, the situation can become even more complex. This is because of the open structure in combination with the flexible connectivity of the constituent units. While the interaction of the single monomer units with the surface corresponds to the one of small molecules at an interface, the covalent linkage between these segments can cause cooperativity and configurational entropic forces. Furthermore, internal segments and end-groups are distinguished in their interaction with the surface, and in the case of copolymers or functional side groups, different segments can have different affinities to adhere and undergo self-assembly. Still for polymer coils and even more for polymer micelles and microgels, one can distinguish inside and outside. Typically the boundary is rather poorly defined and flexible and does not constitute a two-phase dispersion, however, the interaction with a well-defined interface of another phase may be accounted for predominantly by the surface contact and elasticity of the respective soft polymer object.

Hence, it must be considered that surface induced ordering and structure formation in polymer films formed by adsorption or also by deposition of ultrathin polymer films can be controlled simultaneously by all the different effects that are mentioned above as characteristic for the different systems, i.e., small molecules in istropic solution, liquid crystalline liquids as well as hard colloids. In combination with the elastic entropic forces, cooperativity in the interaction of interlinked segments as well as the incorporation of directed bonding elements provide a tremendous variability in the structure and dynamics of interfacial and thin films. The lecture will be directed towards a phenomenological description of different examples with an em-

phasis on the underlying structure property relation and synthetic accessibility rather than a rigorous treatment of the physical background.

In the following we will focus on polymers on flat straight hard surfaces. Firstly, a short summary is given on self-assembled monolayer formation, liquid crystalline ordering and the structure of adsorbed polymers. In the second section the structure formation of ultrathin blockcopolymer films confined in between the surface to the hard substrate and air will be discussed. Emphasis is put on the transition from a bulk like film to molecularly thick films. At the example of brush like macromolecules it will be shown how the molecular conformation can be affected by adsorption/desorption. At the example of microgels we will provoke the question, when do colloids become molecules. Finally we will demonstrate possibilities for defined functional nano-pattern formation by the means of well defined block copolymers.

2 Polymers, Liquid Crystals, and SAMs

Adsorption of low molecular weight compounds from solution is an equilibrium process controlled by the specific affinity of the adsorbed molecules to the respective surface and their molar activity. For linear macromolecules, the situation is more complex. Because the chain segments are linked, adsorption is also hindered by the entropic forces needed to deform the polymer coil. On the other side, the high local concentration of segments and next neighbor effects can cause cooperativity and favor adsorption. As a rule of thumb, it is observed that macromolecules do displace similar low molecular weight molecules. However, because of the entropy loss by chain stretching, tight adsorption in solution enforcing formation of a 2D-conformation is unlikely and the typical structure can be described by formation of loops trains and tails (see Fig.1a) The latter refers to the fact that the endgroups differ in their chemical nature. In the case endgroups can be bound specifically and relatively strongly, formation of end-tethered polymer chains are observed (Fig.1b). Specific endgroup can cause cooperativity is controlled by the equilibrium

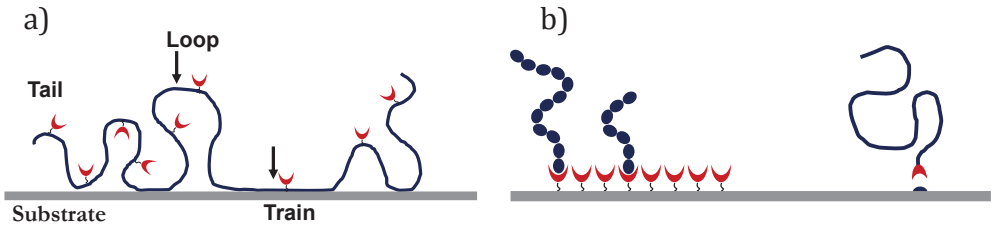


Fig. 1: *a) Adsorption of a polymer chain, left via the segments in the backbone yielding loops, trains and tails; b) selectively via the chain ends yielding end-tethered polymer chains.*

Densely end-grafted polymer chains form a brush-like structure, where the chains get overstretched due to repulsive forces compared to their equilibrium structure (coil dimension) in solution (Fig.2a).

End group adsorption by reversible interaction or weak chemical bonding is typically not sufficient for overcoming the entropic penalty to form a true brush with a thickness exceeding the hydrodynamic radius of the polymer chains in solution, As shown in Fig.2b, a loose brush is formed that allows coiling up of the chains. If the solubility of the molecules that form such a brush is changed from a good to a bad solvent, the chain molecules collapse and can form a mushroom like structure.

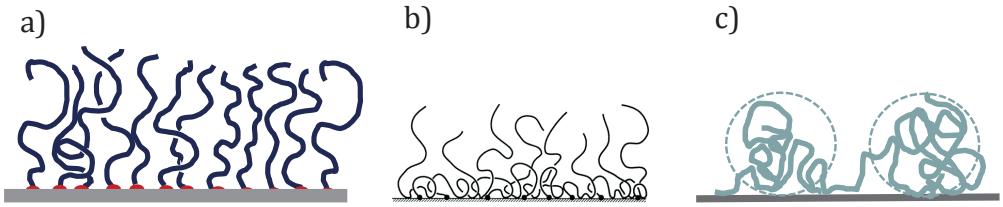


Fig. 2: *a) High density brush with stretched chains, b) low density brush in good solvent and c) low density brush in bad solvent forming a mushroom structure.*

A completely different situation can be encountered in the case of well defined low molecular weight molecules that adsorb via their endgroups but can form an ordered 2D-layer. I this case a self-assembled monolayer, SAM, as depicted in Fig.3 can be formed. Here it is necessary that the distances of the adsorbed head groups, which are controlled by the crystal structure of the substrate match with the lateral packing of the tails. In the case of alkane chains chemisorbed on a gold 111 surface by thiol-end groups, this is achieved by inclination of the alkyl chains.

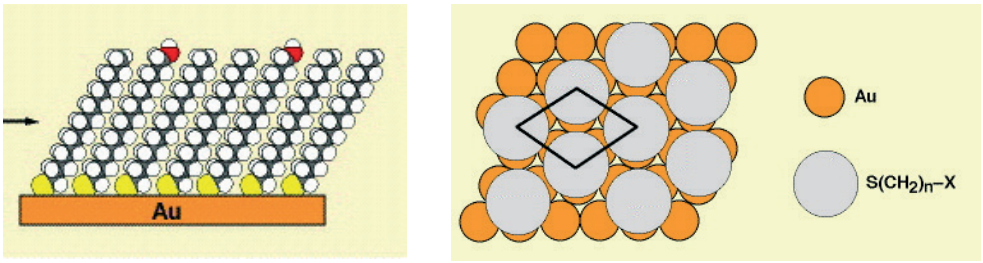


Fig. 3: *Self assembled monolayer of an α-thiolated normal alkane on a gold 111surface. Regular packing of the alkyl chains is enabled by 30° inclination. George M. Whitesides et al., Chem. Rev. 105 (4): 1103–1170, 2005, www.ifm.liu.se/applphys/ftir/sams.html*

A similar situation can be observed even without a specific directed chemisorption, provided the layer structure is sufficiently stable, which is the case for vesicle forming molecules. Fig.4 depicts the principle of a supported bilayer of lipids. In contact with a flat wall, vesicles that are the typical equilibrium structure in aqueous solution can open to form a straight layer that is still separated from the wall by water molecules that solubilize the polar head groups of the lipid.

The latter example demonstrates that molecular orientation is not necessarily limited to the chemisorption, but can be caused by matching of the mesoscopic structures. This is widely

observed in liquid crystals where, e.g., smectic layers typically orient parallel to a flat surface provided directed chemical bonding to the surface is not dominant.

In this case, often a phenomenon can be observed denoted as surface freezing. Even above the melting or isotropization temperature of the molecules that form the liquid in contact with the surface the flat wall can induce the formation of an ordered layer.

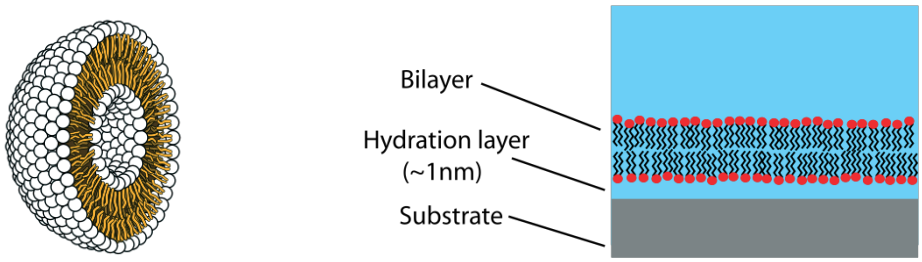


Fig. 4: *Supported lipid bilayer that can be formed when water dispersed lipid or gemini surfactant vesicles get in contact wit a hydrophilic flat surface. http://en.wikipedia.org/wiki/Model_lipid_bilayer*

Thus, the surface layer has a higher melting point than the bulk. Surface freezing is observed, e.g., for normal alkanes and even more pronounced for semifluorinated alkanes. Fig.5 demonstrates this schematically for an alkane, where a segment 12 carbon atoms substituted by fluorine atoms is linked to a hydrocarbon chain with 8-to 12 carbon atoms. It is observed that the melting point of the surface layer can be up to 2-3°C higher than that of the bulk.

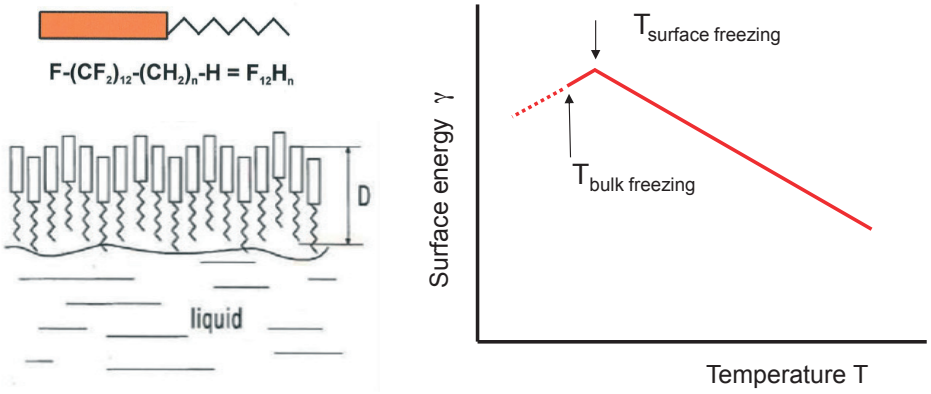


Fig. 5: *Surface freezing of semifluorinated alkanes against air. Supported by the low surface energy of a surface that is formed from densely packed CF_3-endgroups, the first molecular layer organizes in a lamella sheet at temperatures up to 2-3°C above the melting temperature of the bulk. (ref. M. Deutsch, M, Möller et al.)*

The effect can be also observed at the interface of suitable liquid and a solid substrate. Fig.6 depicts an optical micrograph of a thin film of a polyethylenoxide-grafted polysiloxane. Clearly formation of layered structure with the layers oriented parallel to the surface is observed. Here the driving force is the strong phase segregation of the side chains and the poly-

mer backbone. Yet due to the nonuniformity of the molecular structure layer formation is limited to the surface film.

The enormous potential of such effects can be seen at the example of polymethacrylate partially grafted by oligomeric perfluoropropene ether chains. Due to the irregularity of the molecular structure, these polymers cannot crystallize or form a regular bulk structure. Yet, in contact with a flat glass slide or a silicon wafer regular layer formation is observed.

In the case the polyacrylates or polymethacrylates (see Fig.7) are substituted by linear perfluorinated side chains, that can undergo side chain crystallization, the melting temperature of the layers adjacent to the flat substrate has been reported to exceed melting of the bulk structure by up to 40°C.

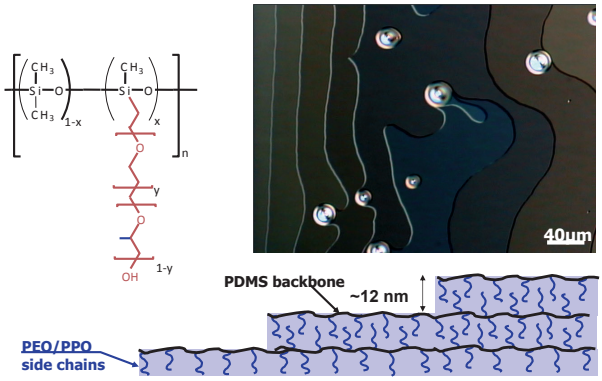


Fig. 6: *Formation of a regular layered structure of a polyether grafted polydimethylsiloxane on a silicon wafer (unpublished result).*

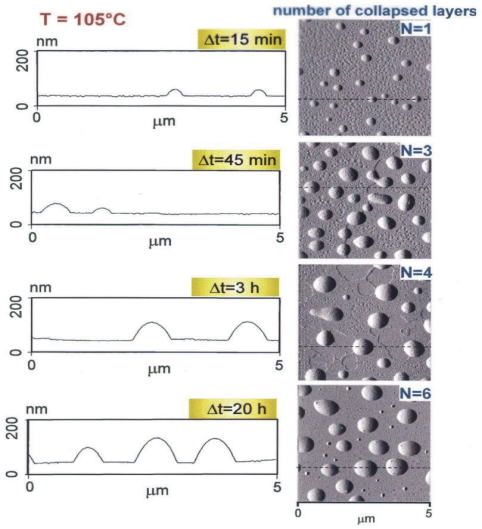


Fig. 7: *Scanning force micrographs of thin multilayer structure of a polymethacrylate partially substituted by perfluorinated ether side chains. The layered structure formed in contact with a flat substrate becomes metastable at elevated temperature and disorders (melts) layer by layer starting from the air interface of the thin film. S. Sheiko, E. Lermann, M. Möller, Langmuir 4015, 12 (1996).*

3 Adsorption/Desorption of Brush-like Macromolecules

An interesting situation occurs when the adsorption of a polymer is very strong, essentially forcing all segments of the macromolecule to be in contact with the surface, so that the chain must adopt a flat 2D-conformation. If an amorphous, coiled polymer is adsorbed, the transition from the 3D-solution to the tightly adsorbed 2D-structure does not change the fundamental coiling. This is, however, different for brush like molecules as depicted in Fig.8.

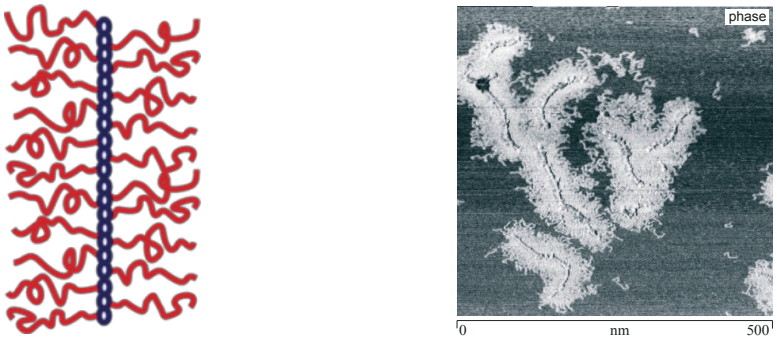


Fig. 8: *(left) Schematical structure of a brush like molecule and (right) a scanning force micrograph of polymethacrylate-graft-Poly(n-butylacrylate) adsorbed on mica (ref. S.S. Sheiko, M. Möller, K. Matyjaszewski et al., Macromolecules 34, 8354 (2001)).*

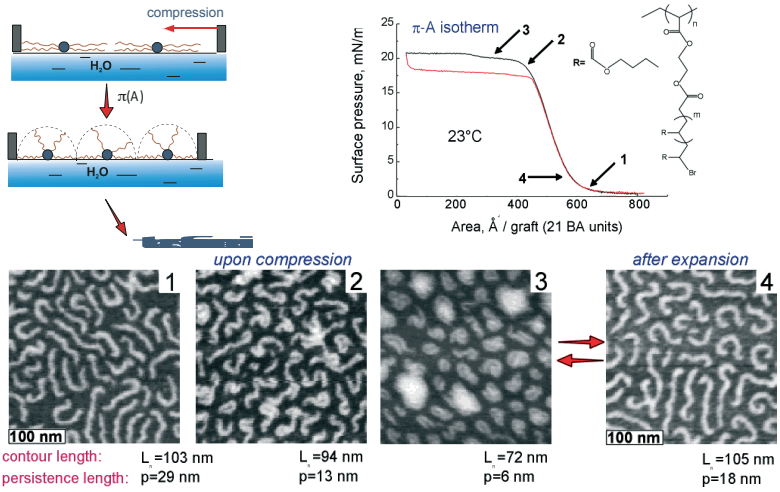


Fig. 9: *Reversible globular collapse of brush molecules adsorbed on water and affected by 2D-compression by means of a Langmuir balance (ref. see Fig.8).*

By the term brush-molecule, we denote a structure where each monomeric unit in the backbone is substituted by a side chain that is sufficiently long to adopt a coil conformation. In solution the side chains can occupy the full space around the backbone, when the side chains, however, are adsorbed tightly on a flat wall their conformational space is reduced significant-

ly and they force the backbone to stretch considerably and the persistence length can be larger than the contour length.

For long side chains, the stress on the backbone that is exerted when the molecules are adsorbed at the surface of water can become so strong that even rupture of the backbone is observed [8]. With side chains of moderate length and on a solid substrate it is possible to observe a transition to single molecule globuli, when the adsorption strength is reduced by co-adsorption of a low molecular weight compound or by lateral compression on a Langmuir balance, as shown in Fig.9.

4 Microgel Adsorption, Molecular-like or Colloid-like

Microgels are ultrahigh molecular mass intra-molecularly cross-linked molecules swollen with a solvent. They are currently investigated due to their role in applications including emulsion stabilizers [9], coatings [10], drug delivery and release [11,] sensors [12], and cell culturing substrates [13]. Highly swollen microgels comprise an open structure with a diffuse outer-boundary, where also the inner segments are well solvated, so that the solute and solvent form a single phase [14]. Thus, the suspended solvent-swollen microgel particles are dissolved like a linear polymer in a good solvent. The high fraction of solvent inside the microgel together with steric stabilization ensures colloidal stability and swelling/deswelling of the elastic network structure. Such peculiarities form a basis for switchability. Due to these distinctions from the hard particles, it may be expected that the interfacial adsorption of microgels demonstrate specific characteristics that are untypical for dispersed hard particles.

In analogy, the interfacial activity of microgels might actually be expected to be less pronounced because of the high solvent fraction in a swollen microgel and the resulting compatibility of the microgel with the solvent [15];[16]. However, a pronounced interfacial activity can be observed which is accompanied by significant deformation of the microgel particle shape.[17] Also at liquid-solid interfaces, adsorption of dispersed nano- and microparticles has been studied intensively [18]. Rigid particles are well known to form highly regular, mostly hexagonally packed monolayer structures. Here the driving forces are colloidal interaction and most importantly capillary forces exerted by the meniscus that evolves upon evaporation of the solvent at the solvent/air/particles interface.[19] Elastic behavior of soft particles can be considered by contact mechanics theory [20] that accounts for the variation of the adhesive contact by balancing adhesion energy, which favors large contact area, against elastic energy, which opposes deformation [21].

To study the effect of the cross-linking concentration on the microgel deformation (spreading) at the air/solid interface, we investigated the adsorption of microgels at the water/air interface by measuring the surface tension of poly(vinylcaprolactam) microgels as a function of time. Fig.10 shows that regardless of the cross-liking degree of the microgel the surface tension decreases and eventually reaches a steady state with an equilibrium value equal to about 46 mN/m. Since the microgel contains 50mol% vinylcaprolactam, we expected the surface tension to be different from previously reported equilibrium surface tension for PNIPAm microgels (~ 42 mN/m) [15]. Remarkably the time required to achieve a steady state of the surface tension increases with the particle cross-linking density. The data demonstrate that weakly cross-linked microgels exhibit marked adsorption rates compared to highly cross-linked ones.

In general, adsorption depends on (i) the concentration (ii) the diffusion rates and in the case of the soft microgels (iii) on the conformational flexibility that allows eventually for some cooperativity in the molecular spreading at the air/water interface. The surface tension data do not differentiate between these effects, but clearly show that more flexible microgels adsorbs faster. The effect that the interfacial activity of the microgel depends on their conformational flexibility is in agreement with the observation that microgels strongly deform upon adsorption at interfaces.

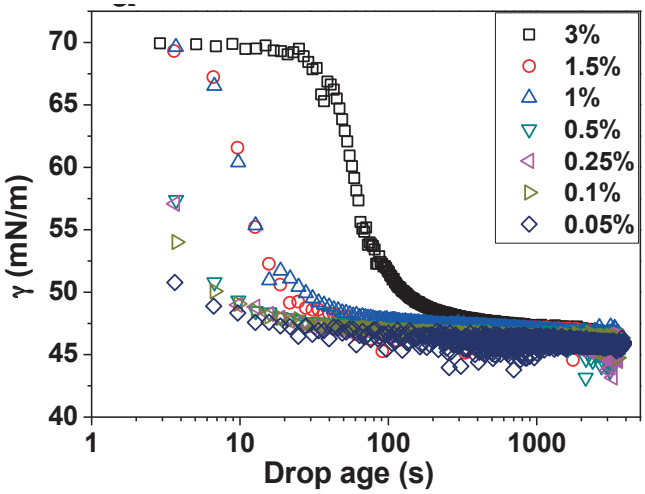


Fig. 10: *Dynamic surface tension for microgels with different cross-linker contents. The microgel concentration was kept constant and corresponds to 0.2 g/L* [22].

In order to investigate the extent of the molecular deformation, i.e., the spreading of the microgels upon interfacial adsorption, we studied the deposited microgels by scanning force microscopy. Single isolated microgels were adsorbed/deposited from dilute solution on a silicon wafer and imaged after solvent evaporation as they were adsorbed.

The experiment is based on the assumption that the monitored molecular structure will indicate whether the segments adsorb from solution or whether the molecules are just deposited as the solvent evaporates. Indeed, the flat and spread shape of the microgel indicates adsorption from solution. Notably, the weakly cross-linked molecules extend on the surface with contact-radius exceeding hydrodynamic radius in solution, as it will be shown below. Because the microgel exhibits a radial gradient in the cross-linker concentration, it is expected that they adopt a conformation, where the softer outer parts are strongly flattened that the densely crosslinked core.

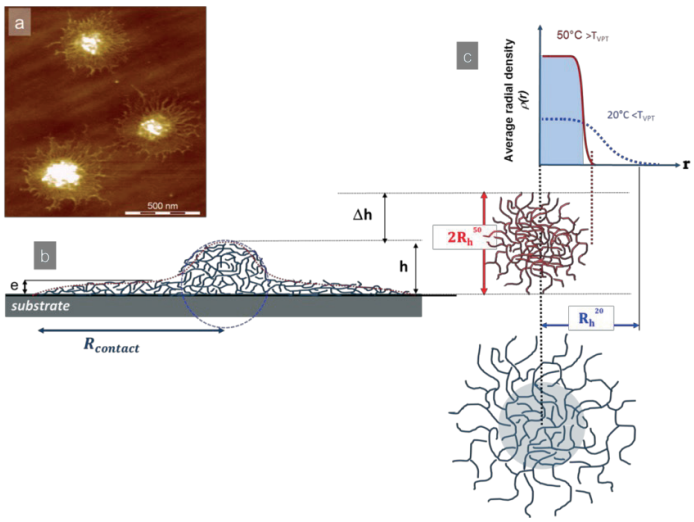


Fig. 11: *(a) Height image of dehydrated microgel on solid surface and (b) a scheme of the height profile and (c) is density profile of a swollen and collapsed microgel -according to reference 14 to highlights the relation between the dense core and height, as well as hydrodynamic radius and contact area* [23].

Fig.11 summarizes the structural parameters of the following discussion. When a compliant colloidal particle with radius R is adsorbed on a flat surface the ratio of the reduced height in the vertical direction Δh/(2R) described the extent of its deformation (Fig.11). A hard sphere does not deform (Δh ~ 0), whereas a highly flexible particles may flatten. In contrast, to an elastic particle that adsorbs via surface contacts, a highly flexible molecule can unfold so that nearly all chain segments get in contact to the surface and the vertical deformation may approach its diameter (Δh ~ 2R). Because of the radial gradient in the cross-linker concentration, the shell and core of a microgel respond differently.

Actually, for microgel adsorbed from solution on a solid surface, we observe a hemispherical core surrounded by a flattened corona of adsorbed polymer chains. The radius of the whole structure is denoted the contact radius, R_{cont}. At the periphery, the height of the corona is molecular and corresponds to the diameter of the constituent polymer chains. The height h of the core is due to the limited deformability of the strongly crosslinked inner segments of the microgel. As a reference for the lateral deformation (spreading) we refer to the hydrodynamic radius of the well-solubilized microgel below the VPT, R_h^{20}; and for the vertical deformation (flattening) to the hydrodynamic radius of the collapsed particle above the VPT represented by R_h^{50}.

In the limit of a particle like behavior where the adsorption is controlled by the surface to surface adhesive contact, we expect mostly the hemispherical shape as depicted for the core. In the case of a loosely crosslinked microgel for which adsorption is controlled also by the inner segments, we expect flat structure as depicted for the corona.

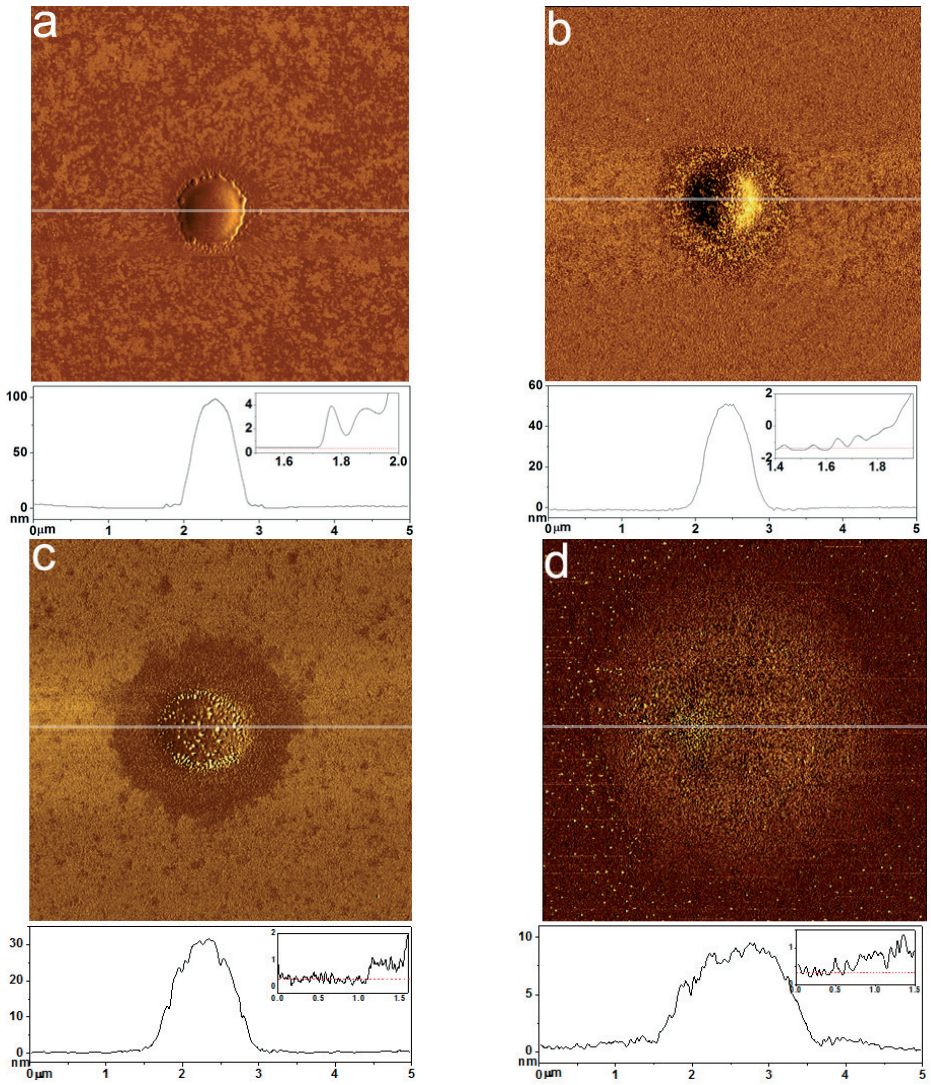


Fig. 12: *Morphology of adsorbed microgels with different cross-linker contents: (a) 3mol%, (b) 1mol%, (c) 0.5mol%, and (d) 0.05mol%. For each image, the scan size is 5×5µm2, and the upper is a phase image and the graph below it is a cross section profile of topographic image at the same position along the horizontal line as indicated. The inset in the height profile highlights the outer shell-like part.*

This is indeed observed for microgels when the crosslinking is varied. In Fig.12a and b, the height profiles for particles with 3mol% and 1mol% crosslinker respectively shows a hemispherical shape. The diameter of the hemisphere is roughly in agreement with the hydrodynamics radius of the well-solubilized microgel. A different observation has been encountered when the crosslinking density has been reduced to 0.5 and 0.05mol% crosslinker. In the first case the adsorbed molecules exhibit a well pronounced “fried egg” shape with an extended molecularly flat (1nm) corona around a hemispherical core (Fig.12c).

In the second case, the molecules are further spread and the hemispherical core is significantly reduced in height. The widespread corona exhibits a height corresponding to tightly adsorbed trains of the polymer segments (1nm). Quantitatively, Fig.13 depicts height (h) of the hemisphere, and the radius of the spread molecules including the corona, i.e., the contact radius (R_{cont}) depending on the crosslinking density. For comparison, we also depicted the hydrodynamic radii of the microgel in solution, under good solvent (at 20°C) as well as under bad solvent conditions (collapsed microgel at 50°C). The graph demonstrates that the profile of microgels on surface and the corresponding dimensions in solution are comparable for highly crosslinked spheres. However, a gradual decrease in the crosslinking density results in extended contact radius concurrently flattens the profile to almost monomer-thickness.

The data indicate that the physical behavior of microgel on surface may be separated in three regimes of dissimilar crosslink density. (1) Spreading of a highly crosslinked microgel is defined by the elastic stresses that develop upon deformation. The influence of the surface tension is reduced to the microgel edge. The physical behavior of the microgel is completely dominated by the elastic properties of the network. (2) Intermediate crosslink state where spreading forces and elastic stresses conjointly defines the shape of the microgel on surface. (3) Loosely crosslink microgel flattens to form a very thin pancake.

The thickness is defined by the long-ranged van der Waals interactions described by the Hamacker constant A, $h \sim (A/S)^{1/2}$. Connectivity and a radial gradient of crosslinker density limit the lateral extension whereby the central part of the adsorbed microgel is slightly thicker. In this regime, the microgel is highly stretched; the physical behavior is dominated by the elastic properties of the individual chains.

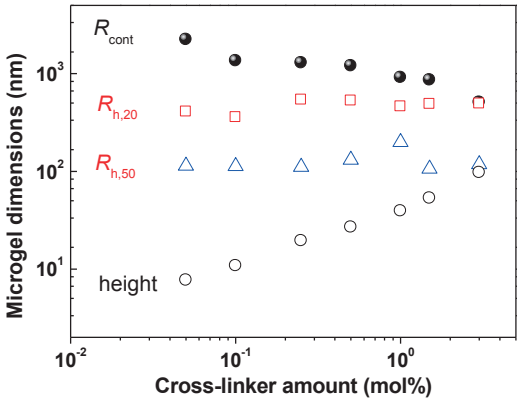


Fig. 13: *Microgel dimensions in contact with the solid surface and in solution depending on the cross-linker content. Its height and contact radius denoted by R_{cont} describes the shape of the microgel on SiO_2. For comparison, the hydrodynamic radius in swollen ($R_{h,20}$) and collapsed state ($R_{h,50}$) [22].*

The large deformation indicates tight adsorption and spreading forces that must be active in the presence of solvent. In particular, a loosely cross-linked microgel sustains large defor-

mation due to swelling **and** adsorption further amplifies the strain and eventually may induce uncontrolled CC-bonds scission leading to partial fragmentation of the microgel (see Fig.14).

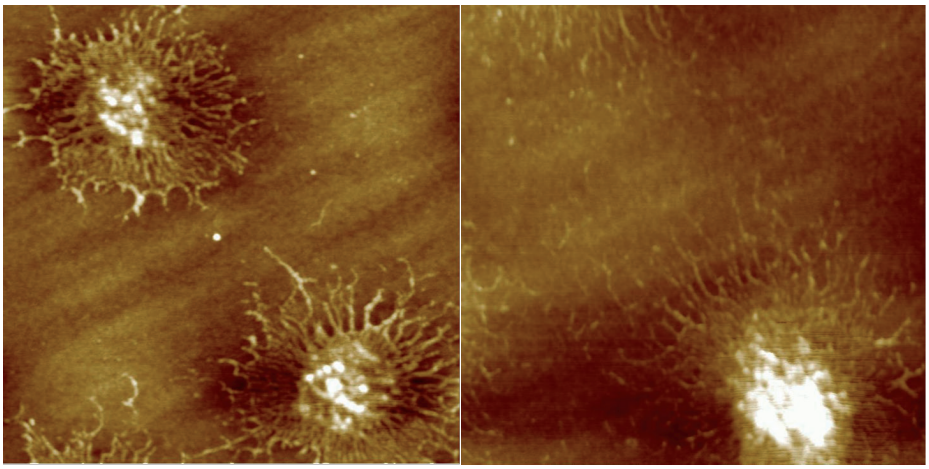


Fig. 14: *SFM Height images of microgel with fragmented polymer strand at their periphery.*

5 References

[1] J. Penfold, Reports on Progress in Physics, 64, 777 (2001)
[2] J.A. Prins, H. Petersen, Physica 3, 147 (1936)
[3] S. Baer Physica 87A, 569 (1979)
[4] P.S. Pershan, Faraday Discuss. Chem. Soc. 89, 231 (1990)
[5] A.S. Dimitrov, Nagayama*, K.,* Langmuir 12*,* 1303 *(*1996*)*
[6] K. Ariga, Y. Yamauchi, T. Mori,J.P. Hill, Advanced Materials 25, 6477 (2013)
[7] J. C. Love, L. A. Estroff, J.K. Kriebel, R. G. Nuzzo, G.M. Whitesides, Chem. Rev. 105, 1103, (2005)
[8] J.C. Love, L.A. Estroff, J.K. Kriebel, R.G. Nuzzo, G.M. Whitesides, Chem. Rev. 105, 1103, (2005)
[9] Brugger, B.; Rosen, B.A.; Richtering, W., Langmuir**,** 24, 12202 (2008); Berger, S.; Zhang, H.P.; Pich, A., Adv. Funct. Mater. 19, 554 (2009); Brugger, B.; Richtering, W., Langmuir 24, 7769 (2008)
[10] Wen, Q.; Pelton, R., Microgel adhesives for wet cellulose: Measurements and modeling. Langmuir, 28, 5450 (2012); Wen, Q.; Vincelli, A.M.; Pelton, R., Cationic polyvinylamine binding to anionic microgels yields kinetically controlled structures. J. Colloid Interface Sci. 369, 223 (2012)
[11] Bysell, H.; Mansson, R.; Hansson, P.; Malmsten, M., Adv. Drug Delivery Rev. 63, (13), 1172 (2011); Wanakule, P.; Liu, G.W.; Fleury, A.T.; Roy, K., Journal of Controlled Release 2012, 162, 429 (2012).
[12] Ancla, C.; Lapeyre, V.; Gosse, I.; Catargi, B.; Ravaine, V., Langmuir, 27, 12693 (2011); Wang, D.; Liu, T.; Yin, J.; Liu, S.Y., Macromolecules, 44, 2282 (2011).
[13] Saunders, B.R.; Laajam, N.; Daly, E.; Teow, S.; Hu, X.H.; Stepto, R., Adv. Colloid Interface Sci. 147-48, 251 (2009).; Schmidt, S.; Zeiser, M.; Hellweg, T.; Duschl, C.; Fery, A.; Mohwald, H.,. Adv. Funct. Mater. 20, 3235 (2010)
[14] Pelton, R., Adv. Colloid Interface Sci. 85, 1 (2000); Heyes, D.M.; Branka, A. C., Soft Matter, 5, 2681 (2009); Stieger, M.; Richtering, W.; Pedersen, J.S.; Lindner, P., 120, (13), 6197 (2004); Lyon, L.A.; Fernandez-Nieves, A., Annual Review of Physical Chemistry, 63, 25 (2012)
[15] Zhang, J.; Pelton, R., Langmuir 15, 8032 (1999).
[16] Binks, B.P., Current Opinion in Colloid & Interface Science*,* 7, 21 (2002).
[17] Destribats, M.; Lapeyre, V.; Wolfs, M.; Sellier, E.; Leal-Calderon, F.; Ravaine, V.; Schmitt, V., Soft Matter, 7, 7689 (2011); Deshmukh, O.S.; Maestro, A.; Duits, M.H.G.; van den Ende, D.; Stuart, M.C.; Mugele, F., Soft Matter 2014; Vasilevskaya, V.V.; Khalatur, P. G.; Khokhlov, A.R., Conformational Polymorphism of Amphiphilic Polymers in a Poor Solvent. Macromolecules 36, 10103 (2003).
[18] Pieranski, P., Physical Review Letters 45, 569 (1980)
[19] Kralchevsky, P.A.; Nagayama, K., Langmuir, 10, 23 (1994)
[20] Johnson, K.L.; Kendall, K.; Roberts, A.D., Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 324, 301 (1971).
[21] Carrillo, J.M.Y.; Raphael, E.; Dobrynin, A.V., Langmuir, 26, 12973 (2010); Salez, T.; Benzaquen, M.; Raphael, E., Soft Matter, 9, (45) 10699 (2013)
[22] From A. Mourran, Y. Wu, R.A. Gumerov, A.A. Rudov, I.I. Potemkim, A. Pich, M. Möller, in preparation.

A 5 Membrane Proteins & Cellular Signaling

Ch. Fahlke
Institute of Complex Systems
Forschungszentrum Jülich GmbH

Contents

1 Introduction 2

2 Organization and function of biological membranes 2

2.1 Transport across biological membranes 2

2.2 Molecular basis of ion transport by channels and transporters 4

3 Electrical signals 7

3.3 Action potentials in signal propagation 9

4 Communication between cells 10

References 12

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

Cells are the functional units of all living organisms. In complex multicellular organisms distinct cells fulfil different tasks, and the differentiation of embryonic cells into specialized cells in the adult body is of major importance for normal body function. Such specialization requires coordination and thus communication between individual cells. In this chapter I will focus on the generation of a special type of cellular signaling, fast electrical signaling, the propagation of such signals within cells and their transmission from one cell to the other.

2 Organization and function of biological membranes

Every living cell is surrounded by a cell membrane that defines the cell as unit, receives and releases signal molecules and thus mediates communication between the intracellular and the extracellular space. Biological membranes are based on a lipid bilayer. This design is of exceptional simplicity and furthermore provides unique mechanical and electrical properties. The fluidity of the lipidic matrix permits cell division and cell movement. A lipid bilayer restricts the permeation of polar and charged molecule and thereby effectively defines intra- and extracellular milieu. Its impermeability for charged molecules makes the lipid bilayer a perfect electrical isolator and represents the basis for electrical signaling.

The electrical capacitance of a lipid bilayer is very low. For a typical size with a diameter of 30 µm, the cell capacitance is only 30 pF. This value signifies that only 2.4 pC, i.e. 1.5 x 10^7 elementary charges need to be moved across the membrane to generate transmembrane voltages of about 80 mV. This amount of ions can be moved across the membrane within one second by a single ion channel. Changing the transmembrane potential is fast and does not require major energy consumption. Electrical signaling therefore dominates fast communication in virtually all cells of the human body.

2.1 Transport across biological membranes

Since the lipid bilayer is impermeable for polar and charged molecules there are specialized membrane transport proteins that mediate the selective transport of such molecules across the membrane. Membrane transport proteins are usually categorized into channels, transporters and pumps (1).

Ion channels permit the fastest transport of ions across the membrane, reaching transport rate of up to more than 10^8 ions per second (1). They are usually involved in fast ion transport in the generation of electrical signals by changing the cell´s membrane potential. Channels differ from transporters and pumps in the simplicity of their transport function (Fig. 1). They are characterized by an aqueous conduction pathway that permits the diffusion of ions from one membrane side to the other. Only passive ion diffusion is possible via ion channels. There are two driving forces for ion diffusion: the concentration gradient and the voltage. These two driving forces can be combined to the electrochemical gradient:

$$\Delta G = RT \ln \frac{c_i}{c_o} + zFV \qquad (1)$$

with R being the gas constant, T the absolute temperature, c_i and c_o: intra- and extracellular ion concentration, z the charge of the ion, F the Faraday constant, and V the transmembrane voltage.

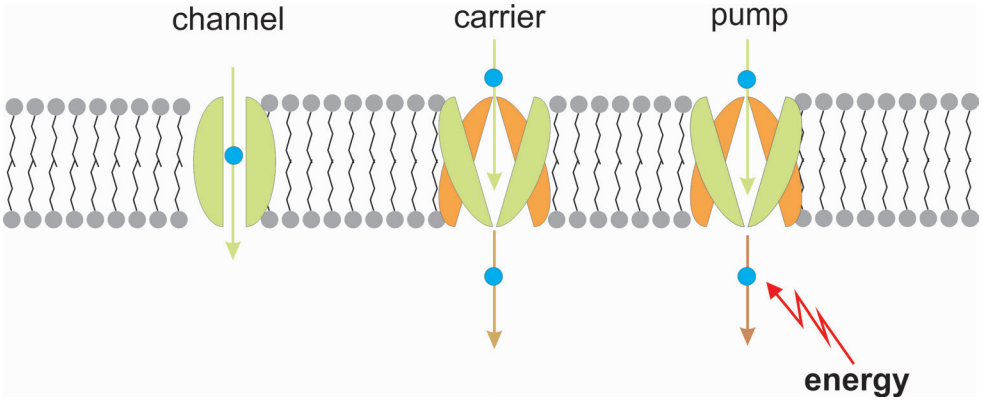


Fig. 1: *Schematic drawing of a lipid bilayer with three types of membrane transport proteins.*

Ion channels conduct ions in one direction at negative electrochemical potentials and in the other direction at positive values. At an electrochemical potential of 0, ion efflux and influx are identical resulting in zero net ion transport.

Transporters and pumps differ from ion channels in the mechanisms underlying transport. In these two groups of membrane proteins substrates are moved across the membrane via a conformational change of the protein. These changes can be discrete, for example by providing alternating access of substrates to or from a central binding site to the internal and external space by changing the direction of only few side chains. Other transport processes are based on the movement of major components of the transporter protein across the membrane (see below). Since conformational changes of proteins occur at much lower rates than diffusion processes transport rates of transporter are much lower than those of ion channels.

There are passive transporters that move substrates following the electrochemical gradient across the membrane. Secondary-active transporters transport more than one ion: the transport of one substrate occurs passively, and the transport along its electrochemical gradient provides the energy to move other substrates against their electrochemical gradients. According to the relative transport directions of the substrates symporters and antiporters are distinguished. Symporters move all substrates in the same direction, whereas antiporters mediate the simultaneous import of one substrate and export of the other substrate. Secondary-active transporters normally function at a fixed transport stoichiometry: a certain number of substrate A is only transported in co- or countertransport with a certain number of substrate B.

Ion pumps are specialized transporters that can move substrates against their electrochemical gradient using energy provided by ATP hydrolysis.

2.2 Molecular basis of ion transport by channels and transporters

For many decades ion channels and transporters could only be studied on a functional level. Detailed structural insights that are the prerequisite for understanding transport processes at near atomic resolution have not existed until 15 years ago. The first high-resolution structure of a membrane transport protein was solved from a bacterial potassium-selective ion channel, KcsA from *Streptomyces lividans* (2). This structure beautifully explained the basis for selective ion conduction through biological membrane (Fig. 2).

As I will describe in more detail in the following sections, electrical signaling in mammalian cells is based on regulated membrane permeabilities for the two main cations in biological media, Na^+ and K^+. The plasma membranes of resting cells are usually Na^+ impermeant, but highly conductive for K^+. Cell excitation is based on the transient opening of Na^+-permeant channels. These processes thus require ion channels that reliably select between these two ions. This important task is complicated by the fact that Na^+ and K^+ carry identical charges and only differ minimally in their diameter (Na^+: 1.9 Å, K^+ 2.6 Å).

High resolution structures of K^+-selective channels demonstrate how these requirements can be fulfilled (Fig. 2). KcsA exhibits a short narrow constriction, the selectivity filter, in which ions can only enter after complete dehydration. K^+ channels are tetrameric and each subunit carries a short glycine-tyrosine-glycine (GYG) sequence in the narrow part of the selectivity filter. The carbonyl oxygens of the tyrosine side chain and of the inner glycine perfectly substitute for interactions of the K^+ ions with water molecules. The smaller Na^+ ions interact less favourably with the selectivity filter and are thus excluded from entering the channel.

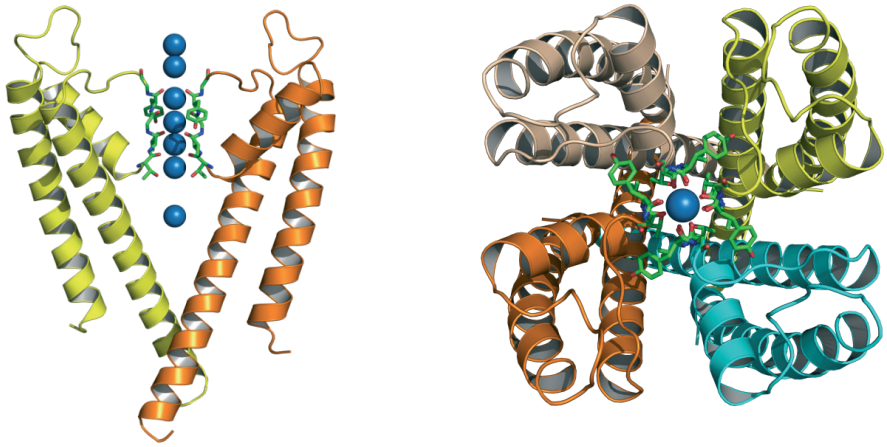


Fig. 2: *Backbone fold of KcsA shown from the side (left panel) or from outside (right panel). Blue circles depict permeating K^+ ions that are either entering or leaving or bound to the GYG motif within the selectivity filter. In the side view two of the four subunits of the tetrameric assembly are removed for better illustration.*

This design permits very tight association of K^+ ions within the channel protein. Since effective electrical signaling requires high transport rates, selective ion channels do not only have to bind ions with high affinity, but also to permit rapid binding and unbinding. In K^+-selective channels tight binding is overcome by the occupation of the selectivity filter with multiple K^+ ions. K^+ ions that bind within the selectivity filter coming from outside or onside

the cell destabilize binding of the other ions and permit the release of the outermost ion to the opposite side of the membrane. The enormous complexity of generating a K^+-selective ion channel is illustrated by the fact that only one mechanism for K^+ selectivity has been developed during evolution. In contrast, there are multiple protein families and designs for channels selective for Na^+, Ca^{2+} and Cl^- ions.

Secondary-active transporters function by a complete different mechanism as ion channels. Whereas ions permeate through an aqueous conduction pathways of ion channels, ion transporters move ions across the membrane utilizing conformational changes. To understand these conformation changes crystal structures of such transporters have to be solved in multiple conformations. In the following I will illustrate secondary-active transport on one particular example, Na^+-coupled glutamate transport (Fig. 3, 4 and 5). There are multiple other transporter families that function using alternative mechanisms.

Glutamate is the major excitatory neurotransmitter in the mammalian central nervous system. After release from presynaptic nerve terminals, glutamate diffuses across the synaptic cleft and binds to specialized postsynaptic receptors. Activation of these receptors results in the generation of electrical and/or chemical signals in the postsynaptic cells. Glutamatergic synaptic transmission is terminated by the uptake of glutamate into surrounding glial and neuronal cells. This transport is the main physiological task of Na^+-coupled glutamate transporters, the SLC1 family (3). Coupling glutamate and Na^+ transport permits the establishment of low extracellular glutamate concentrations resulting in improved sensitivity of glutamatergic synaptic transmission.

High resolution crystal structures of a bacterial glutamate transporter from *Pyrococcus horikoshii*, Glt_{Ph}, in multiple conformations (4-8) illustrated how these transport processes occur and how movement of diverse substrates are coupled. Glt_{Ph} is assembled as trimer from three identical subunits that function independently of each other (Fig. 3). Each subunit exhibits eight transmembrane helices with two hairpin loops (HP1 and HP2) that open and close during transport.

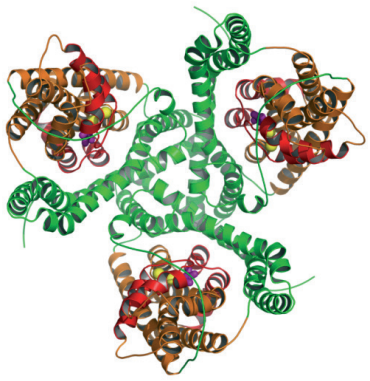


Fig. 3: *Backbone fold of Glt_{Ph} shown from the outside. Glt_{Ph} is a trimer assembled from three identical subunits. In each subunit the static trimerization domain (green) is distinguished from the mobile transport domain (orange) with HP1 (yellow) and HP2 (red).*

Transport is initiated by binding of the first Na^+ ion to the apo structure (Fig. 4). This binding step causes HP2 (pink in Fig. 4) to open permitting association of glutamate/aspartate and the remaining Na^+ ions. After closure of HP2 a major conformation change occurs, the translocation of the transport domain (orange in Fig. 3, yellow in Fig. 4) across the

membrane. This translocation represents the ~18 Å rotational-translational movement of the transport domain relative to the membrane. After release of the substrates in the inward-facing conformation the transporter returns via re-translocation to its original conformation.

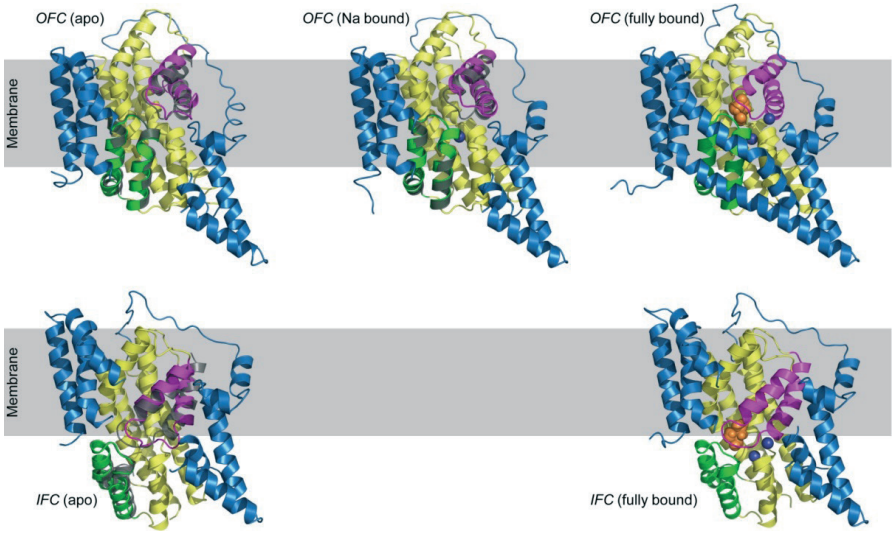


Fig. 4: *Conformational transitions of a single Glt_{Ph} subunit during coupled amino acid transport.*

Although transport occur at much lower rates in secondary-active transporters than in channels, transporters have to solve the same conflicting tasks of a selective channel, selection among different substrates and effective transport. Glt_{Ph} uses an induced fit mechanisms to select between different amino acids (9). Transport substrates bind after opening of HP2 first loosely to the transporter. The subsequent closure of HP2 prevents dissociation of the substrate and is necessary for the transport of the substrate across the membrane.

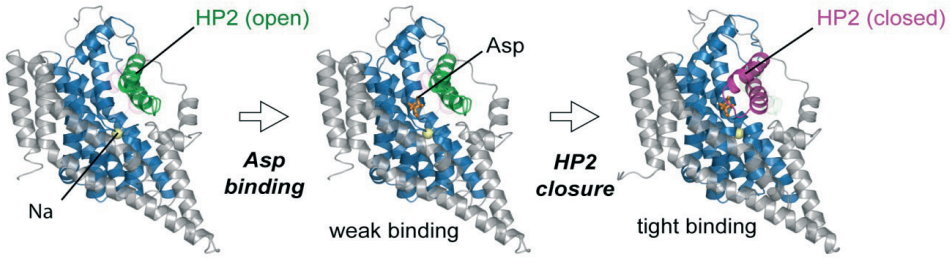


Fig. 5: *Induced fit mechanisms of substrate selection illustrated for an individual Glt_{Ph} subunit.*

At the other membrane side opening of the other hairpin loop then permits release of the substrate into the cytoplasm. The two processes involved in substrate association, initial binding and hairpin loop closure permits selection of the main substrate from other substrates that bind initially tighter to the transport domain and can therefore not be transported at high rates.

3 Electrical signals

3.1 The resting potential

Every cell exhibits a voltage across the plasma membrane and across most intracellular membrane systems. When measured in cells that are not generating electrical signals this voltage is usually denoted as resting potential. I will use in the following mostly this physically incorrect term, since it is the preferred term within physiology and biophysics. Measuring these voltages requires electrical access to the inside of the cell without major impairment of the electrical isolation by the membrane, for example by impaling a fine glass electrode into the cell. For many cells, for example skeletal muscle or glial cells, one can observe a voltage of about -80 mV after impalement and short regeneration of the membrane. In these cells, the membrane potential critically depends on the extracellular potassium concentration. Plotting the resting membrane potential of a skeletal muscle fibre one observes a dependence that can be almost perfectly described by the Nernst equation (10). This equation describes diffusion potentials at the border between solutions with different concentrations. From this mathematical description it was concluded that the cellular resting potential is - in many cells - a potassium diffusion potential and does not directly depend on energy consuming processes.

In the human body intracellular K^+ concentrations are about 150 mM, due to the high density of primary active Na^+-K^+-ATPases. The extracellular [K^+] is tightly regulated to low levels, with concentrations between 3.6 and 5.6 mM. Under these conditions, cells with a plasma membrane that is selectively permeant to K^+ ions will exhibit a K^+ diffusion potential that is close to the experimentally determined values.

One can develop the generation of a K^+ diffusion potential of these values in a model cell that only exhibits K^+-selective ion channels and physiological internal and external [K^+]. In this model cell, K^+ will initially diffuse out of the membrane driven by the concentration gradient. Since only K^+ can cross the membrane K^+ efflux will result in charge separation and in the development of a transmembrane voltage. Net K^+ efflux will stop when the chemical driving and the electrical driving force compensate each other so that the electrochemical gradient for K^+ ion is zero. At this moment net K^+ flux across the membrane stops and ion concentrations as well as the membrane potential are stable.

$$\Delta G = RT \ln \frac{c_i}{c_o} + zFV \tag{2}$$

resulting in the Nernst equation

$$V = \frac{RT}{zF} \ln \frac{c_o}{c_i} \tag{3}$$

3.2 The action potential

Excitable cells are capable of generating action potentials, transient changes of the cellular membrane potential in response to a depolarizing stimulus. A prerequisite for action potential generation is the abundant existence of voltage-gated sodium channels in certain regions of the cell (11). These sodium channels exhibit unique gating properties: they are closed at negative potentials and open upon cell depolarization with steep voltage dependence. After reaching the so-called threshold potential few voltage-gated sodium channels open. Mammalian cells exhibit small intracellular [Na^+] so that opening of sodium channel results in passive influx of Na^+ into the cells and in further depolarization of the cells. The particular gating properties of the voltage-dependent sodium channels results in reinforcing sodium influx (Fig. 6). Changes of the membrane potential to more positive values cause opening of additional sodium channels making the action potential a self-energizing process that will result - once initiated - always in similar electrical signals.

Sodium currents are terminated by an additional conformation change of these sodium channels, the inactivated state in which sodium channels are closed and unable to re-open. Only the return of the membrane potential to negative values permit transitions from the inactivated to the closed states from which the channels can open again upon depolarization of the threshold potential. Sodium channel inactivation initiates the return of the action potential to the resting potential. The delayed opening of voltage-gated potassium channels that permit K+ efflux and return to the K+ diffusion potential enhances membrane repolarization.

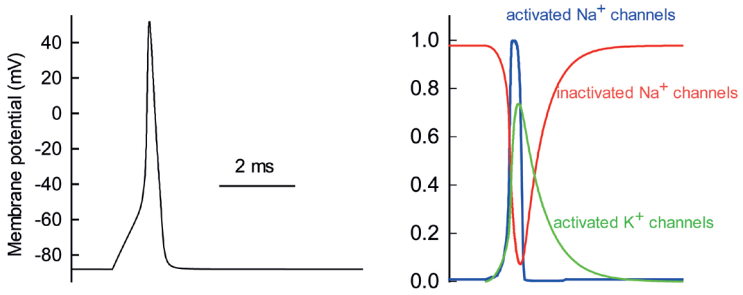


Fig. 6: *Idealized action neuronal action potential. The action potential is initiated by an extracellular signal that depolarizes the cell (initiation phase). After reaching the threshold potential Na^+ channels open, and Na^+ ion entering into the cell result in further depolarization of the membrane potential (action potential upstroke). Na^+ channel inactivation and opening of voltage-gated K^+ channels result in membrane repolarization.*

The steep voltage dependence of sodium channels results in the generation of similar action potentials by a given type of cell independently of the stimulus. This so-called “all-or-nothing” rule of action potential has been known for more than 100 years.

Inactivation of sodium channels prevents the generation of an action potential immediately after the end of the proceeding signal. Only after the refractory period action potentials can be again elicited (Fig. 7). One can distinguish an absolute refractory period in which no action potential can be elicited and a relative refractory period with only reduced action potential amplitudes.

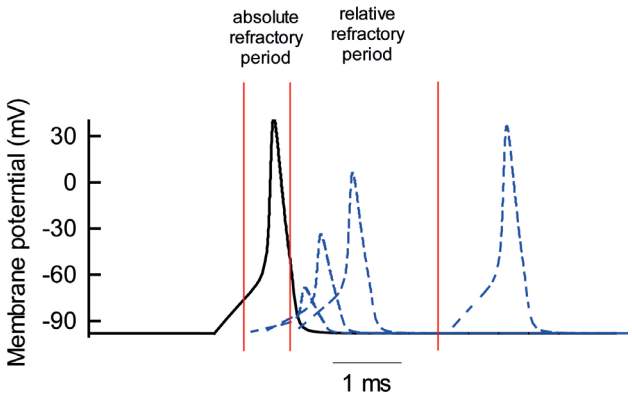


Fig. 7: *Refractory periods after action potential initiation. Immediately after an action potential no action potential or only action potentials with reduced amplitudes can be elicited by external stimuli.*

3.3 Action potentials in signal propagation

An important role of action potential generation is the propagation of electrical signals over long distances in cells with complex morphology. Over small distances electrical signals are propagated in a passive manner, similar to electrical propagation in electrical cables. However, due to the particular electrical properties of cells such propagation is very ineffective and results in a rapid decline of signal amplitudes within only short distances.

The length constant is a characteristic parameter that quantifies the effectivity of passive electrotonic propagation of electrical signals. Within one length constant the amplitude of an electrical signal diminishes to 1/e. Local changes in membrane potential result in ion currents within the cytoplasm, across the membrane and outside the cell. Whereas a high resistance of the cell membrane (R_m) prevents current flow and thus results in a reduced decline of electrical signal at a certain distance of the signal, cytoplasmic (R_i) and extracellular (R_o) resistances – that determine the voltage drop by this particular current amplitude – have opposite effects on the length constant. This reasoning provides a qualitative explanation for the dependence of the membrane length constant λ on these three resistances.

$$\lambda = \sqrt{\frac{R_m}{R_i + R_o}} \tag{4}$$

In the human body high intracellular resistances or low membrane resistances result in length constants of only few µm so that passive propagation will not allow propagation of electrical signals in axons or muscle fibres with lengths up to 1 m. Action potential generation allows

regeneration of the electrical signal during propagation over long distances. Within one length constant an action potential depolarizes neighbouring membrane areas above the threshold potential. Opening of sodium channels results in generation of novel action potentials in the membrane areas so that the initial electrical signal is regenerated. By this mechanism an electrical signal can be propagated without decrease in amplitude over long distances.

Action potential generation also permits separation of electrical signals from electrical noise. Only signals that are capable of eliciting action potentials will be propagated whereas signals below threshold further decrease and disappear within few length constants.

Regenerative propagation of electrical signals by action potentials illustrates the physiological importance of the refractory period at the end of the action potential. Refractory periods prevent back propagation of the electrical signals into the direction of its origin. Sodium channel inactivation is thus a necessary mechanism to ensure directed propagation of electrical signals.

In the human body many peripheral nerves exhibit a myelin sheet. This myelin sheet accelerates the propagation of electrical signals by multiple mechanisms (1). Myelin sheets are generated by specialized glial cells, Schwann cells, which wrap the neuronal extension with multiple lipid layers originating from the original cell membrane. The myelin sheet increases the length constant and thus decreases the number of action potentials necessary for signal propagation along a given distance. Since each action potential is an energy-consuming process myelinisation reduces the energy demand of electrical signalling. The most important effect of the myelin sheet is the reduction of the cell capacitance. Reduced cell capacitances decrease the number of ions necessary for particular electrical signals and thus greatly accelerate the generation of electrical signals.

4 Communication between cells

Cells communicate at specialized contact structure synapses. Two functionally distinct types of synapses can be distinguished by their different mode of function, electrical and chemical synapses.

Chemical synapses are characterized by gap-junction channels, high diameter ion channels that cross the cell membrane of both cells that form the synapse. Gap junction channels are not only permeable for ions and thus allow fast transmission of electrical signals from one cell to the other one, but also the diffusion of second messenger. Electrical synapses therefore electrically and chemically couple cells. They play important roles in the electrical synchronization of certain types of smooth muscle, of cardiomyocytes in the heart and of glial cells.

Compared with electrical synapses, the architecture of chemical synapses is very complex. Excitation of presynaptic cells causes the release of neurotransmitters via exocytosis of synaptic vesicles. In synaptic vesicles neurotransmitter are accumulated by secondary-active transporters to very high concentrations. Multiple synaptic proteins in the synaptic vesicles and in the membrane of the synaptic nerve terminal enable fast and reliable fusion of the vesicles if and only if the presynaptic nerve terminal is excited. After crossing the synaptic cleft neurotransmitters bind to postsynaptic receptors. There are two classes of postsynaptic neurotransmitter receptors: ionotropic and metabotropic receptors. Activation of ionotropic receptors results in the opening of cation- or anion-selective ion channels. Influx of cations causes depolarization of the postsynaptic cell, resulting in the excitatory postsynaptic

potentials. In inhibitory synapses, anion currents results in membrane hyperpolarization and/or increased resting cell conductances and length constants.

The higher physiological importance of chemical synapses despite their complex architecture is based on several unique properties of such synapses. Chemical synapses permit excitation as well as inhibition of the postsynaptic cell. They only transmit information in one direction, from the pre- to the postsynaptic cell and thus direct the propagation of signals through the central nervous system. Lastly, synaptic transmission in chemical synapses can be regulated by multiple mechanisms. Synaptic plasticity as basis of higher brain functions such as learning and memory are mainly based on such regulatory pathways affecting pre- or postsynaptic processes.

References

[1] B. Hille, *Ion Channels of Excitable Membranes*, 3 ed., Sinauer Associates Inc., Sunderland, MA, (2001)
[2] D.A. Doyle, J.M. Cabra, R. Pfuetzner, et al., *Science* 280, 69 (1998)
[3] N. Danbolt, *Progr Neurobiol* 65, 1 (2001)
[4] D. Yernool, O. Boudker, Y. Jin, E. Gouaux, *Nature* 431, 811 (2004)
[5] O. Boudker, R.M. Ryan, D. Yernol, et al., *Nature* 445, 387 (2007)
[6] N. Reyes, C. Ginter, O. Boudker, *Nature* 462, 880 (2009)
[7] G. Verdon, S. Oh, R.N. Serio, O. Boudker, *eLife* 3, e02283
[8] S. Jensen, A. Guskov, S. Rempel, et al., *Nat Struct Mol Biol* 20, 1224 (2013)
[9] D. Ewers, T. Becher, J.P. Machtens, et al., *Proc Natl Acad Sci USA* 110, 12486 (2013)
[10] A.L. Hodgkin, P. Horowicz, *J Physiol (Lond)* 148, 127 (1959)
[11] A.L. Hodgkin, A.F. Huxley, *J Physiol (Lond)* 117, 500 (1952)

THEORY

B

B 1 Dynamics of Colloids

Jan K.G. Dhont
Institute of Complex Systems / ICS-3: Soft Condensed Matter
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**
1.1 Probability density functions (pdfs) . . . 3
1.2 Overdamped dynamics and force- and torque balance . . . 3

2 Translational Diffusion **4**
2.1 Fick's diffusion equation . . . 5
2.2 The continuity cquation . . . 8
2.3 Einstein's expression for the diffusion coefficient, and the Brownian force . . . 9
2.4 Translation diffusion of non-spherical colloids . . . 9

3 Rotational Diffusion **10**
3.1 The orientational diffusion equation . . . 11
3.2 The orientational continuity equation . . . 13
3.3 The rotational diffusion coefficient, and the Brownian torque . . . 14

4 Applications: Combined Translational and Rotational Diffusion in the Presence of an Electric Field, Active Motion, and Shear Flow **15**
4.1 Electric permittivity . . . 16
4.2 Diffusion and active motion of a spherical microswimmer . . . 19
4.3 Diffusion in shear flow: Taylor dispersion . . . 22

5 Diffusion of Interacting Colloids **24**
5.1 The Smoluchowski equation . . . 24
5.2 The concentration dependence of Fick's diffusion coefficient . . . 25
5.3 A shear-induced instability . . . 29
5.3.1 The advection-diffusion equation . . . 29
5.3.2 The equation of motion for the suspension flow velocity . . . 33
5.3.3 Stability and flow profiles . . . 34

A The Contraction Notation **37**

B Gauss's and Stokes's Integral Theorems and Partial Integration **37**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Colloidal particles are objects in the size range of a few nano-meter up to about ten microns, embedded in a fluid. The lower size limit is set by the requirement that the colloids are much larger than the fluid molecules, so that the fluid can be regarded as a structureless continuum. Interactions between colloids mediated through the fluid can thus be described on the basis of macroscopic equations of motion for fluid flow. These interactions are therefore referred to as hydrodynamic interactions. In addition to these fluid-mediated hydrodynamic interactions, colloids interact through direct forces, which are due to mutual exclusion of volume, and, for example, charges on the surface of the particles and/or polymers that are chemically attached to the colloid's surfaces. The upper size limit is set by the requirement that the colloidal particles exhibit vivid thermal motion. When a particle in a fluid is relatively large, the asymmetry in the typical number of collisions with fluid molecules from two opposite sides of the particle is small, resulting in a small net instantaneous force that sets the particle in motion. The motion due to collisions with fluid molecules is nothing but thermal motion that every molecule experiences. Larger particles thus exhibit weaker thermal motion, which ceases to occur for typical sizes of a few tens of microns. Colloidal particles are thus nothing but "large molecules", which exhibit thermal motion that drives self-assembly, phase transitions, and dynamics, similar to atomic/molecular systems.

Contrary to atomic/molecular systems, where inter-particle interactions are set by the rules of quantum mechanics, the direct interactions between colloids can be tuned almost at will. Interactions between colloids can be tuned by chemical surface functionalization of the colloidal particles, by adding small molecules to the fluid (so-called depletion interactions), or by changing the quality of the solvent (resulting in van der Waals interactions). These types of interactions will be discussed in more detail in **lecture B4**. Of course these interactions are also subjected to the rules of quantum mechanics, but the composition of the colloidal particles can be changed, contrary to atomic/molecular systems. In addition, the geometrical form of the core of the colloids can vary from spherical, to plate-like and rod-like. This freedom leads to a large class of materials with many practical applications. According to the above definition of colloids, biological macromolecules belong to this class of systems, ranging from proteins up to entire cells and bacteria. Colloid science can thus contribute to a quantitative understanding of living systems.

Due to their large size, thermal motion can be observed through light microscopy. It was first observed by Ingenhousz in 1785, and later in 1827 by Brown, although at that time the cause of the erratic motion was not at all understood. Thermal motion of colloidal particles is nowadays commonly referred as Brownian motion, and colloidal particles are some times referred to as Brownian particles. Brownian motion is at the origin of so-called diffusive mass transport, to which this chapter is devoted. Translational diffusion relates to the motion of the center-of-mass of a colloidal particle, while rotational diffusion relates to changes of orientation of colloidal particles due to the same interactions with fluid molecules as mentioned above. We discuss combined translational and rotational diffusion, also in the presence of electric fields, shear flow, and active motion.

After introducing probability density functions in section 1.1 and the notion of overdamped dynamics in section 1.2, translational diffusion is addressed in section 2, and rotational diffusion in section 3. In section 4 a few applications of combined translational and rotational diffusion are discussed: the dielectric response of polar molecules to an oscillating electric field, diffusion of active swimmers, and diffusion in shear flow. This relates so far to non-interacting

colloids. Section 5 deals with more concentrated systems, where interactions between the colloidal particles play an important role. The effect of interactions on diffusion are described, and a shear-induced instability is analyzed, where interactions are shown to give rise to mass transport from regions of high shear rate to low shear rate, leading to an instability that gives rise to a non-uniform flow profiles.
Some of text books on diffusion and microstructural order where most of the material of this lecture can be found are [1, 2, 3, 4, 5].

1.1 Probability density functions (pdfs)

A deterministic description involving positions and velocities of single particles in a macroscopic system is not possible, as this requires the solution of Newton's equations for a very large number of interacting particles. Newton's equation of motion can be solved through computer simulations, but analytical theory is not feasible. Instead, a theory can be developed in terms of *probabilities* of events, like an instantaneous displacement of a colloidal particle due to collisions with a large number of fluid molecules. The theory based on probability concepts in physics is referred to as *statistical mechanics*. Here, probabilities are quantified by so-called *probability density functions* (pdfs). Consider a variable X, like for example the center-of-mass position of a colloidal particle. The probability density function (pdf) $P(X,t)$ is defined, such that $P(X,t)\,dX$ is the probability, at time t, that the variable attains a value within the interval $(X, X+dX)$, with dX a (infinitesimally) small increment of X. The average of any function $f(X)$ is thus equal to,

$$< f > (t) \;=\; \int dX\; f(X)\, P(X,t)\;. \tag{1}$$

This average is the experimentally accessible quantity, where many independent realizations of X are probed. It can be time dependent when the system is out-of equilibrium, through the time dependence of the pdf. In the sequel we will encounter vectorial analogues of X, like the position coordinate and the orientation of a colloidal particle.

1.2 Overdamped dynamics and force- and torque balance

Equations of motion for probability density functions will be derived for the translational and orientational Brownian motion of colloidal particles on the so-called *diffusive time scale*, where inertial forces can be neglected. The diffusive time scale is defined as follows. Consider a spherical colloidal particle, embedded in a fluid, with an initial translational velocity $\mathbf{v}_0$. In the absence of forces due to interactions with other colloidal particle or an external field, the only force acting on the sphere is the friction force with the fluid, which is equal to $-\zeta\,\mathbf{v}(t)$, where $\zeta > 0$ is *the friction coefficient*, and $\mathbf{v}(t)$ the thermally averaged velocity at time t. This is the velocity averaged over many random collisions with solvent molecules. The thermally averaged velocity satisfies Newton's law of motion, which states that the inertial force (= mass $\times$ acceleration) is equal to the total force acting on a particle,

$$M\,\frac{d\mathbf{v}(t)}{d\,t} \;=\; -\,\zeta\,\mathbf{v}(t)\;,$$

where M is the mass of the particle. The solution to this equation of motion is,

$$\mathbf{v}(t) \;=\; \mathbf{v}_0\,\exp\{-\,t/\tau\} \quad , \quad \text{with} \quad \tau \;=\; M/\zeta\;.$$

As the friction coefficient for a sphere with radius R is equal to $\zeta = 6\pi\eta_0 R$, with η_0 the shear-viscosity of the fluid, typical values for the time constant τ, for radii of colloidal particles in the nano-meter to micron size range, are found to be of the order of a nanosecond. Colloidal particles thus loose their velocity due to friction with the solvent in about a nano-second. A very similar analysis based on Newton's equation of motion for the orientation of colloidal particles shows that the same holds for rotational motion. In all cases where the interest is in the dynamics on much longer time scales than a nano-second, inertial forces and torques may therefore be neglected. The dynamics of colloids on this longer time scale is referred to as *overdamped dynamics*, or *diffusive dynamics*. The time scale $\tau_D \gg M/\zeta$ on which neglect of inertial forces is allowed defines *the diffusive time scale*. In this lecture we restrict ourselves to such diffusive dynamics. In the overdamped limit, according to Newton's equation of motion, the total force $\mathbf{F}^{tot}$ and torque $\mathbf{T}^{tot}$ on a colloidal particle (averaged over a time interval larger than a nano-second) are thus zero,

$$\mathbf{F}^{tot} \;=\; \mathbf{0} \quad , \quad \mathbf{T}^{tot} \;=\; \mathbf{0}\,. \tag{2}$$

This equation expresses *force- and torque-balance on the diffusive time scale*, which will be important later in the derivation of diffusion equations.

2 Translational Diffusion

The main interest in this section is concerned with the probability density function (pdf) $P(\mathbf{r},t)$ of the center-of-mass position $\mathbf{r} = (x,y,z)$ of a single spherical colloidal particle, at time t. This pdf is defined, such that $P(\mathbf{r},t)\,dx\,dy\,dz$ is the probability, at time t, that the center-of-mass is located within the small volume element $dx\,dy\,dz$ that is located at (x,y,z). The variable X in eq.(1) is now the vector $\mathbf{r} = (x,y,z)$. Instead of integration with respect to just a single variable X in eq.(1), we now have a three-fold integration (or "a volume integration") with respect to the x-, y-, and z-coordinates of the center-of-mass position ,

$$< f > (t) \;=\; \int dx \int dy \int dz\; f(\mathbf{r})\, P(\mathbf{r},t) \;\equiv\; \int d\mathbf{r}\; f(\mathbf{r})\, P(\mathbf{r},t)\;,$$

where in the second equation $\int d\mathbf{r}$ is nothing but a short-hand notation for the integral in the first equation, with $d\mathbf{r} = dx\,dy\,dz$ an infinitesimally small volume element.

In subsection 2.1, we derive an equation of motion for the pdf $P(\mathbf{r},t)$ which is known as *Fick's diffusion equation* (named after Adolf Fick, who formulated this equation of motion in the 1850's), and discuss its fundamental solution. In subsection 2.2 the so-called *continuity equation* is derived, which is an exact equation of motion for the pdf, but contains the averaged velocity of the Brownian particle as an unknown quantity. In order to generalize Fick's diffusion equation to include external fields, the Brownian force is identified in subsection 2.3 by comparing Fick's equation to the continuity equation, and employing force-balance as quantified in eq.(2). The so-called *diffusion coefficient* that appears in Fick's diffusion equation is identified in subsection 2.3 in terms of the friction coefficient of the Brownian particle with the fluid. Finally, in subsection 2.4, the extension of Fick's diffusion equation to a non-spherical particle (with fixed orientation) with a cylindrical symmetry is discussed.

2.1 Fick's diffusion equation

Imagine a collection of (infinitely) many, macroscopically identical systems consisting of a single Brownian particle embedded in a fluid. "Macroscopically identical" means that the systems contain the same number of fluid particles with the same temperature and the same volume. Such a collection of macroscopically identical systems is commonly referred to as an *ensemble*, and can be envisioned as a stack of these systems as depicted in Fig.1a (left panel). The ensemble is itself isolated from the surroundings, that is, there is no exchange of heat, volume, or mass possible between the ensemble and the surroundings. The systems can exchange heat among each other (so that they have the same temperature), while their volume is fixed and the number of fluid molecules is fixed. Such an ensemble is referred to as an NVT-ensemble ("N", "V", and "T" standing for the prescribed number of particles, volume, and temperature, respectively).
Although the systems are macroscopically identical, they are generally "microscopically different", in the sense that positions and velocities of the fluid molecules and the Brownian particle are different. Consider now the following *gedanken* experiment, where a photo is made of the entire ensemble at a particular time t. For each system in the ensemble one can determine the position coordinate of the Brownian particle at time t. Collecting all position coordinates at time t gives rise to a distribution, as depicted in Fig.1a (right panel), where the local density of points is proportional to $P(\mathbf{r},t)$. The density-distribution generally changes with time, as the Brownian particles exhibit thermal motion.
Let $W(\mathbf{r} \mid \mathbf{r}')\,d\mathbf{r}'$ denote the probability, per unit time, that a Brownian particle jumps from position $\mathbf{r}$ to a small volume $d\mathbf{r}'$ centered around $\mathbf{r}'$. The temporal change of "the number of points" $P(\mathbf{r},t)\,d\mathbf{r}$ of the density-distribution mentioned above within the volume element $d\mathbf{r}$ is due to jumps inwards and outwards of the volume $d\mathbf{r}$. Since all the points $P(\mathbf{r}',t)\,d\mathbf{r}'$ in $d\mathbf{r}'$ have the probability $W(\mathbf{r}' \mid \mathbf{r})\,d\mathbf{r}$ to jump per unit time into the volume $d\mathbf{r}$, for any position $\mathbf{r}'$ outside $d\mathbf{r}$, the temporal increase of the probability for a particle to reside in $d\mathbf{r}$ equal to,

$$\frac{\partial}{\partial t}\,[\,P(\mathbf{r},t)\,d\mathbf{r}\,]_{(+)} \;=\; \int [\,P(\mathbf{r}',t)\,d\mathbf{r}'\,]\;[\,W(\mathbf{r}'\,|\,\mathbf{r})\,d\mathbf{r}\,]\;,$$

where the volume integral with respect to $\mathbf{r}'$ ranges over the entire space. The number of jumps outwards of the volume element $d\mathbf{r}$ centered at $\mathbf{r}$ is equal to the number of points $P(\mathbf{r},t)\,d\mathbf{r}$ in the volume $d\mathbf{r}$, multiplied by the probability $W(\mathbf{r} \mid \mathbf{r}'\,d\mathbf{r}')$ to jump into any of the volume elements exterior to $d\mathbf{r}$. The temporal decrease of the probability to reside in $d\mathbf{r}$ is thus equal to,

$$\frac{\partial}{\partial t}\,[\,P(\mathbf{r},t)\,d\mathbf{r}\,]_{(-)} \;=\; -\,[\,P(\mathbf{r},t)\,d\mathbf{r}\,]\int [\,W(\mathbf{r}\,|\,\mathbf{r}')\,d\mathbf{r}'\,] \;=\; -\,P(\mathbf{r},t)\,d\mathbf{r}\;.$$

In the second equation it is used that,

$$\int d\mathbf{r}'\;W(\mathbf{r}\,|\,\mathbf{r}') \;=\; 1\;, \tag{3}$$

which simply expresses that the probability for a particle to jump to a position is unity. The following equation of motion for the pdf is thus found,

$$\frac{\partial P(\mathbf{r},t)}{\partial t} \;=\; \int d\mathbf{r}'\;P(\mathbf{r}',t)\,W(\mathbf{r}'\,|\,\mathbf{r}) - P(\mathbf{r},t)\;. \tag{4}$$

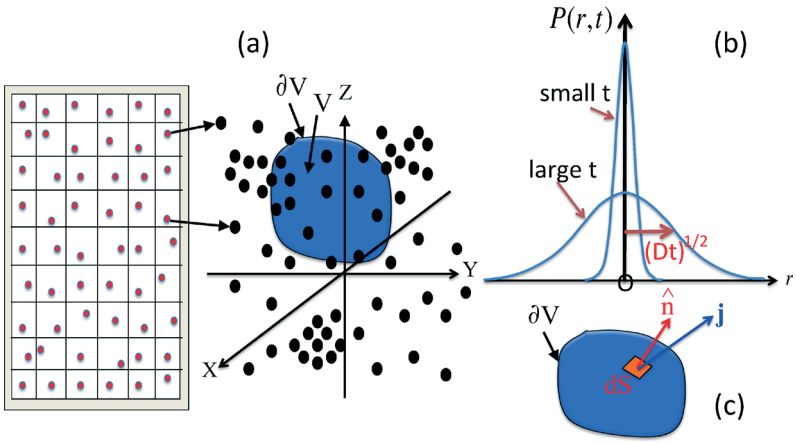


Fig. 1: *(a) The NVT-ensemble (left), which is a stack of macroscopically identical systems containing a single Brownian particle (in red), and a fixed number of fluid molecules (not shown). The collection of instantaneous position coordinates (right) of the Brownian particle in each of the systems in the ensemble. The two arrows are examples for position coordinates for two arbitrary chosen systems. (b) The pdf in eq.(10) as a function of the distance r from the origin, for two times. (c) The rate-of-change of "the number of points" in the volume V, indicated in blue, is due to in- and out-flux of points through the boundary ∂V. dS is an infinitesimally small surface element on ∂V, while $\hat{\mathbf{n}}$ is a unit vector, locally perpendicular to the surface, pointing outwards from the volume V.*

The assumption now is that the transition probability $W(\mathbf{r}\,|\,\mathbf{r}')$ is sharply peaked around $\mathbf{r}' = \mathbf{r}$, as compared to the distances over which the pdf varies. The pdf in the integrand can then be Taylor expanded up to second order (the notation used here is discussed in appendix A),

$$P(\mathbf{r}',t) \;=\; P(\mathbf{r},t) + (\mathbf{r}'-\mathbf{r})\cdot\nabla P(\mathbf{r},t) + \tfrac{1}{2}\,(\mathbf{r}'-\mathbf{r})(\mathbf{r}'-\mathbf{r}) : \nabla\nabla P(\mathbf{r},t)\;, \tag{5}$$

where *the nabla operator* or *gradient operator* is a vector operator equal to,

$$\nabla \;=\; \left(\frac{\partial}{\partial x}\,,\,\frac{\partial}{\partial y}\,,\,\frac{\partial}{\partial z}\right)\;. \tag{6}$$

The Taylor expansion (5) is nothing but the three-dimensional generalization of the well-known Taylor expansion of functions of just one variable. Substitution of this Taylor expansion in the above equation of motion gives (with $\mathbf{R} = \mathbf{r}' - \mathbf{r}$),

$$\frac{\partial P(\mathbf{r},t)}{\partial\, t} \;=\; [\,\nabla P(\mathbf{r},t)\,]\cdot\int d\mathbf{R}\;\mathbf{R}\,W(R) + [\,\nabla\nabla P(\mathbf{r},t)\,] : \int d\mathbf{R}\;\mathbf{RR}\,W(R)\;.$$

where we denoted $W(R) \equiv W(\mathbf{r}'\,|\,\mathbf{r})$, since the transition probability only depends on $\mathbf{r}'$ and $\mathbf{r}$ through the length $R = \mid \mathbf{r}'-\mathbf{r} \mid$ between the two points. The first integral on the right hand-side is zero by symmetry, while the second integral can be written as,

$$\int d\mathbf{R}\;\mathbf{RR}\,W(R) \;=\; \oint d\hat{R}\;\hat{\mathbf{R}}\hat{\mathbf{R}}\times\int_0^\infty dR\;R^4\,W(R) \;=\; \frac{4\pi}{3}\,\hat{\mathbf{I}}\times\int_0^\infty dR\;R^4\,W(R)\;,$$

where $\hat{\mathbf{R}} = \mathbf{R}/R$ is the unit vector in the direction of $\mathbf{R}$, while the first integral ranges over all directions of $\hat{\mathbf{R}}$, that is, it is a surface integral ranging over the unit-spherical surface. Furthermore, $\hat{\mathbf{I}}$ is the three-dimensional identity matrix. We thus finally obtain *the diffusion equation*,

$$\frac{\partial P(\mathbf{r},t)}{\partial t} \;=\; D\,\nabla^2 P(\mathbf{r},t)\;, \tag{7}$$

where the so-called *Laplace operator* ∇^2 is a short-hand notation for,

$$\nabla^2(\cdots) \;=\; \nabla\cdot\nabla(\cdots) \;=\; \left[\,\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\,\right](\cdots)\;,$$

and where,

$$D \;=\; \frac{2\pi}{3}\int_0^\infty dR\;R^4\,W(R)\;, \tag{8}$$

is referred to as *the diffusion coefficient*. Since the transition probability has the dimension $m^{-3}s^{-1}$, the dimension of the diffusion coefficient is m^2/s.

Since the probability to find a particle at a given position is proportional to the local concentration, the diffusion equation also describes the temporal evolution of concentration. It was Fick who put the equation of motion (7) for the concentration of colloidal particles forward for the first time, so that this equation of motion is commonly referred to as *Fick's diffusion equation*.

Consider a particle that is at the origin at time $t = 0$, and subsequently exerts Brownian motion. The average position is equal to zero, as displacements in all directions are equally probable. The most simple average that characterizes Brownian motion is the so-called *mean-squared displacement* $W(t) =< r^2 >$ (with $r =\mid \mathbf{r} \mid$ the distance of the center-of-mass from the origin). $W(t)$ can be calculated by first multiplying eq.(7) on both sides with r^2 and then integrate,

$$\frac{dW(t)}{d\,t} \;=\; \int d\mathbf{r}\;r^2\,\frac{\partial P(\mathbf{r},t)}{\partial\,t} \;=\; D\int d\mathbf{r}\;r^2\,\nabla^2 P(\mathbf{r},t) \;=\; D\int d\mathbf{r}\;P(\mathbf{r},t)\,\nabla^2 r^2 \;=\; 6\,D\;,$$

where two partial integrations have been performed (see appendix B). Since $W(t=0)=0$, this implies that,

$$W(t) \;=\; 6\,D\,t\;. \tag{9}$$

This linear time dependence of the mean-squared displacement is typical for diffusive motion. The solution of eq.(7), under the condition that the particle is at the origin at time $t = 0$, is,

$$P(\mathbf{r},t) \;=\; \frac{1}{(\,4\,\pi\,D\,t\,)^{3/2}}\;\exp\left\{-\frac{x^2+y^2+z^2}{4\,D\,t}\right\}\;. \tag{10}$$

This is a bell-shaped function of the distance $r =\mid \mathbf{r} \mid= \sqrt{x^2+y^2+z^2}$, as sketched in Fig.1b, with a width equal to $\sqrt{Dt}$. The width increases with time, as the particle is more likely to make excursions over larger distances.

2.2 The continuity equation

Here we derive the so-called *continuity equation*, which is an exact equation of motion for $P(\mathbf{r},t)$ in terms of a probability flux. This equation allows to derive an explicit expression for the diffusion coefficient, and to extend Fick's law (7) to include, for example, external fields and active motion of the Brownian particle.
Consider "the number of points" $N(t) = \int_V d\mathbf{r}\ P(\mathbf{r},t)$ in an arbitrary volume V (indicated in blue in Fig.1c). The rate-of-change $dN(t)/dt$ is equal to,

$$\frac{d\,N(t))}{d\,t} \ = \ \int_V d\mathbf{r}\ \frac{\partial}{\partial\, t} P(\mathbf{r},t)\ . \tag{11}$$

On the other hand, this rate-of-change is determined by the in- and out-flux of particles through the boundary ∂V of the volume V. To quantify the in- and out-flux, the flux $\mathbf{j}(\mathbf{r},t)$ is defined as the vector with a magnitude that is equal the number of points that pass a unit surface area per unit time that is perpendicular to the direction of $\mathbf{j}$, and which direction is along the local thermally averaged velocity $\mathbf{v}$ of the particles. By definition we thus have,

$$\mathbf{j}(\mathbf{r},t) \ = \ \mathbf{v}(\mathbf{r},t)\, P(\mathbf{r},t)\ . \tag{12}$$

The component of the flux parallel to a local surface element on the boundary ∂V does not lead to a change of the number of points in V. Only the perpendicular component gives rise to an in- and out-flux. Let $\hat{\mathbf{n}}$ denote the unit-vector (a vector with length unity) that is locally perpendicular to the boundary (see Fig.1c). The component along the perpendicular direction to ∂V is equal to $\mathbf{j}\cdot\hat{\mathbf{n}}$. This is the number of points that enter or leave the volume V per unit time and unit surface area. Summing over all local surface elements dS thus gives,

$$\frac{d\,N(t)}{d\,t} \ = \ \oint_{\partial V} dS\ \hat{\mathbf{n}}(\mathbf{r})\cdot\mathbf{j}(\mathbf{r},t)\ .$$

The surface integral can be cast into a volume integral with *Gauss's integral theorem* (Gauss's theorem is given in appendix B). Equating the result to the expression in eq.(11) gives,

$$\int_V d\mathbf{r}\ \left[\frac{\partial}{\partial\, t} P(\mathbf{r},t) + \nabla\cdot\mathbf{j}(\mathbf{r},t)\right] \ = \ 0\ .$$

Since this holds for arbitrary volumes V, the integrand must be zero, and hence,

$$\frac{\partial}{\partial\, t}\, P(\mathbf{r},t) \ = \ -\,\nabla\cdot\mathbf{j}(\mathbf{r},t) \ = \ -\,\nabla\cdot\left[\,\mathbf{v}(\mathbf{r},t)\, P(\mathbf{r},t)\,\right]\ , \tag{13}$$

where eq.(12) for the flux has been used. Note that this *continuity equation* is an exact equation. Approximations come in when expressing the averaged velocity in terms of the pdf $P(\mathbf{r},t)$ in order to obtain a closed equation of motion.
Fick's diffusion equation (7) is found from the above continuity equation by assuming that, [1]

$$\mathbf{j}(\mathbf{r},t) \ = \ -D\,\nabla P(\mathbf{r},t)\ , \tag{14}$$

which is sometimes referred to as *Fick's first law*. This assumption was originally used by Fick to derive his diffusion equation (7), which is sometimes referred to as *Fick's second law*. It can be shown that Fick's first law (14) for the flux is only valid when gradients in concentration are not too large. When the concentration significantly varies over distances of the order of the size of the colloids, higher order derivatives of the concentration enter the flux equation.

[1] Note that $\nabla f(\mathbf{r})$ is a vector, for any function f, with components $(\partial f/\partial x,\ \partial f/\partial y,\ \partial f/\partial z)$; see eq.(6).

2.3 Einstein's expression for the diffusion coefficient, and the Brownian force

As discussed in section 1.2, inertial forces are negligible as compared to the friction force of the particle with the fluid. The total force is thus zero (see eq.(2)). The friction force is equal to $-\zeta\,\mathbf{v}$, where $\zeta > 0$ is referred to as the *friction coefficient*. Let $\mathbf{F}$ denote the remaining, non-frictional and non-inertial forces. According to eq.(2), $\mathbf{F}^{tot} = \mathbf{F} - \zeta\,\mathbf{v} = \mathbf{0}$. The average velocity of a particle is thus equal to $\mathbf{v} = \mathbf{F}/\zeta$. The continuity equation (13) can thus also be written as,

$$\frac{\partial}{\partial t}\,P(\mathbf{r},t) \;=\; -\frac{1}{\zeta}\,\nabla\cdot[\,P(\mathbf{r},t)\,\mathbf{F}(\mathbf{r},t)\,]\;. \tag{15}$$

Comparing this with Fick's law (7), it is found that $\mathbf{F} \;=\; -\,(\zeta\,D\,)\;\nabla\ln P$ (this force will be discussed further at the end of this subsection). If there is an additional force through the action of an external field, this force has to be added to obtain the total force acting on the particle. Suppose that there is a gravitational field acting in the $-\,z$-direction. giving rise to a force $-\,m\,g$, with m the mass of the particle and g the gravitational acceleration. The total force in eq.(15) is thus $\mathbf{F}(z,t) \;=\; -\,(\zeta\,D\,)\;\nabla\ln P(z,t) - m\,g$. Substitution into eq.(15) gives, noting that P now only depends on z,

$$\frac{\partial}{\partial t}\,P(z,t) \;=\; \frac{\partial}{\partial z}\left[\,D\,\frac{\partial}{\partial z}P(z,t) + \frac{m\,g}{\zeta}\,P(z,t)\right]\;.$$

It follows that in the stationary state (where $\partial P(z,t)/\partial z = 0$), we have: $D\partial P/\partial z + (mg/\zeta)\,P = 0$, and hence $P \sim \exp\{-mgz/\zeta D\}$. On the other hand we know from equilibrium statistical mechanics that $P \sim \exp\{-mgz/k_BT\}$ (with k_B Boltzmann's constant, and T the temperature). This formula is a special case of a "Boltzmann exponent", and is also known as the barometric height formula. It describes the diminished concentration (of gas molecules) at large heights. Comparing the two results immediately leads to *Einstein's formula for the diffusion coefficient*,

$$D \;=\; \frac{k_B\,T}{\zeta}\;. \tag{16}$$

This result was derived independently by Sutherland in 1904, before Einstein's publication in 1905, and later by von Smoluchowski in 1906.
The force $\mathbf{F} \;=\; -\,(\zeta\,D\,)\;\nabla\ln P$ that we mentioned above, in the absence of the gravitational field (or any other field), is referred to as *the Brownian force*, and will from now on be denoted as $\mathbf{F}^{Br}$. From Einstein's formula (16) the Brownian force is found to be equal to,

$$\mathbf{F}^{Br} \;=\; -\,k_BT\,\nabla\ln P(\mathbf{r},t)\;. \tag{17}$$

This is a force of an entropic nature (as signified by the logarithm of P), and drives the system towards equilibrium, even when particles do not mutually interact with each other through, for example, charges or excluded volume.

2.4 Translation diffusion of non-spherical colloids

The above analysis assumes spherical particles. In particular, the friction coefficient in eq.(16) is assumed independent of the orientation of the particle. For rod-like or plate-like particles,

however, the friction coefficient depends on the orientation of the rod/plate relative to it's velocity. The orientation of a rod is characterized by a unit vector $\hat{\mathbf{u}}$ along it's long axis ("unit vector" means a vector of length unity, as indicated by the hat; see Fig.2a). When the rod moves through the fluid along $\hat{\mathbf{u}}$, that is, along it's long axis, the friction coefficient $\zeta_{\parallel}$ is different from the friction coefficient $\zeta_{\perp}$ when the rods moves in a direction perpendicular to $\hat{\mathbf{u}}$. A velocity $\mathbf{v}$ with arbitrary direction can be decomposed in it's components $\mathbf{v}_{\parallel}$ parallel to its long axis and it's perpendicular component $\mathbf{v}_{\perp}$. The parallel component is equal to (see appendix A on the notation convention used here) $\mathbf{v}_{\parallel} = \hat{\mathbf{u}}\hat{\mathbf{u}} \cdot \mathbf{v}$, while the perpendicular component is equal to $\mathbf{v}_{\perp} = \mathbf{v} - \mathbf{v}_{\parallel} = [\,\hat{\mathbf{I}} - \hat{\mathbf{u}}\hat{\mathbf{u}}\,] \cdot \mathbf{v}$ (where $\hat{\mathbf{I}}$ is the identity tensor). The friction force $\mathbf{F}^h$ (where the superscript "h" stands for "hydrodynamic") is thus equal to,[2]

$$\mathbf{F}^h \;=\; -\zeta_{\parallel}\,\mathbf{v}_{\parallel} - \zeta_{\perp}\,\mathbf{v}_{\perp} \;=\; -\Big\{\,\zeta_{\parallel}\,\hat{\mathbf{u}}\,\hat{\mathbf{u}} + \zeta_{\perp}\,\Big[\,\hat{\mathbf{I}} - \hat{\mathbf{u}}\hat{\mathbf{u}}\,\Big]\,\Big\} \cdot \mathbf{v}\;.$$

According to the force-balance principle discussed in section 1.2, which states that $\mathbf{F}^h + \mathbf{F}^{Br} = \mathbf{0}$, a tensor inversion leads to,[3]

$$\mathbf{v} \;=\; -\mathbf{D}(\hat{\mathbf{u}}) \cdot \nabla \ln P(\mathbf{r},t)\;,$$

with,

$$\mathbf{D}(\hat{\mathbf{u}}) \;=\; D_{\parallel}\,\hat{\mathbf{u}}\hat{\mathbf{u}} + D_{\perp}\,\Big[\,\hat{\mathbf{I}} - \hat{\mathbf{u}}\hat{\mathbf{u}}\,\Big]\;,$$

where,

$$D_{\parallel} \;=\; \frac{k_BT}{\zeta_{\parallel}} \qquad , \qquad D_{\perp} \;=\; \frac{k_BT}{\zeta_{\perp}}\;.$$

This is a generalization of Einstein's expression (16) for the diffusion coefficient of a rod-like (or plate-like) particle. According to the continuity equation (13) we thus finally arrive at the diffusion equation for translational motion of a non-spherical particle (see appendix A for the notation used here),

$$\frac{\partial}{\partial t}P(\mathbf{r},t) \;=\; \nabla \cdot \mathbf{D}(\hat{\mathbf{u}}) \cdot \nabla P(\mathbf{r},t)\;,$$

which is the non-spherical equivalent of Fick's law (7). The orientation $\hat{\mathbf{u}}$, however, is also exhibiting Brownian motion, which is referred to as *rotational Brownian motion*. In order to analyze the motion of a rod we therefore need a similar diffusion equation for the orientation of the rod. Rotational diffusion is the subject of the following section.

3 Rotational Diffusion

For non-spherical macromolecules it is not only the center-of-mass that exhibits random Brownian motion due to collisions with fluid molecules, but also it's orientation changes as a result

[2] The superposition of friction forces relies on the low Reynolds number hydrodynamics that applies to colloidal systems, as is discussed in **lecture B3**.

[3] Using that $(\hat{\mathbf{u}}\hat{\mathbf{u}}) \cdot (\hat{\mathbf{u}}\hat{\mathbf{u}}) = \hat{\mathbf{u}}\hat{\mathbf{u}}$ and $(\hat{\mathbf{u}}\hat{\mathbf{u}}) \cdot (\hat{\mathbf{I}} - \hat{\mathbf{u}}\hat{\mathbf{u}}) = \mathbf{0}$, it is easily shown that the inverse of $\Big\{\zeta_{\parallel}\,\hat{\mathbf{u}}\hat{\mathbf{u}} + \zeta_{\perp}\,\Big[\,\hat{\mathbf{I}} - \hat{\mathbf{u}}\hat{\mathbf{u}}\,\Big]\Big\}$ is equal to $\Big\{(1/\zeta_{\parallel})\,\hat{\mathbf{u}}\hat{\mathbf{u}} + (1/\zeta_{\perp})\,\Big[\,\hat{\mathbf{I}} - \hat{\mathbf{u}}\hat{\mathbf{u}}\,\Big]\Big\}$.

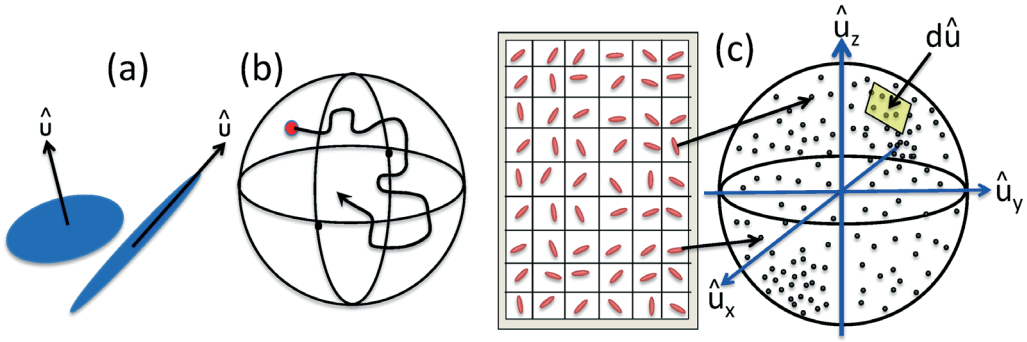


Fig. 2: *(a) A platelet (left) and rod-like (right) colloidal particle, where the unit vector* $\hat{\mathbf{u}}$ *defines their orientation. (b) The tip of the unit vector* $\hat{\mathbf{u}}$*, in red, lies on a unit-spherical surface (the surface of a sphere with radius unity). The red "point-particle" exerts Brownian motion on the unit-spherical surface. (c) An ensemble of macroscopically identical systems containing a single rod in a fluid (left). In each system the rod has a different orientation, which is mimicked as different points on the unit-spherical surface (right). The yellow (infinitesimally small) surface area element has a surface area denoted by* $d\hat{u}$*.*

of these interactions. The erratic change of the orientation is referred to as *rotational Brownian motion*, while the motion of the center-of-mass is some times referred to as *translational Brownian motion*.

We consider a uniaxial Brownian particle, the orientation of which is specified by a unit-vector $\hat{\mathbf{u}}$ ("unit-vector" means that the length of the vector is unity, as indicated by the hat $\hat{}$). For a rod-like particle this vector lies along the long axis of the rod, while for a platelet this vector can be chosen to be perpendicular to the face of the platelet (see Fig.2a). The tip of the unit-vector $\hat{\mathbf{u}}$ lies on a unit-spherical surface. Changes in orientation leads to motion of the tip over the unit-spherical surface. Orientational diffusion can thus be interpreted as the diffusion of a point particle (the tip of $\hat{\mathbf{u}}$) on a unit-spherical surface (as depicted in Fig.2b). Just like for translational motion, we consider an ensemble of macroscopically identical systems, containing a single particle (see Fig.2c, left panel). Instead of the center-of-mass, the interest is now in the orientation $\hat{\mathbf{u}}$ of the particles. The orientation of the particle in each of the systems is now a point on the unit-spherical surface, so that the entire ensemble defines a point distribution as shown in the right panel of Fig.2c. We define the probability density function (pdf) $P(\hat{\mathbf{u}}, t)$ such that $P(\hat{\mathbf{u}}, t)\, d\hat{u}$ is the probability that $\hat{\mathbf{u}}$ takes a value within the small surface area element $d\hat{u}$ on the unit-spherical surface (see Fig.2c.) The point density on the unit-spherical surface in Fig.2c is proportional to the local numerical value of the pdf.

In the sequel we discuss: (i) the derivation of the equation of motion for the probability density function (pdf) for the orientation, (ii) the derivation of a rotational-continuity equation, and (iii) the identification of the diffusion coefficient and the Brownian torque, quite similar to what has been done in section 2 for translational diffusion.

3.1 The orientational diffusion equation

Let $W(\hat{\mathbf{u}} \mid \hat{\mathbf{u}}')\, d\hat{u}'$ be the probability that the orientation jumps, per unit time, from $\hat{\mathbf{u}}$ to a $\hat{\mathbf{u}}'$ within the surface area element $d\hat{u}'$ (as depicted in Fig.3a). The *orientational transition proba-*

bility $W(\hat{\mathbf{u}}\,|\,\hat{\mathbf{u}}')$ has units of $1/s$. The arguments leading to eq.(4) for translational diffusion can be copied to give the similar equation of motion for rotational motion,

$$\frac{\partial P(\hat{\mathbf{u}},t)}{\partial\, t} \;=\; \oint d\hat{u}'\; P(\hat{\mathbf{u}}',t)\, W(\hat{\mathbf{u}}'\,|\,\hat{\mathbf{u}}) - P(\hat{\mathbf{u}},t)\;.$$

The transition probability is again assumed to be sharply peaked in comparison to the variation of the pdf, so that the pdf in the integrand can be Taylor expanded up to second order (for an explanation of notation, see appendix A),

$$P(\hat{\mathbf{u}}',t) \;=\; P(\hat{\mathbf{u}},t) + (\hat{\mathbf{u}}'-\hat{\mathbf{u}})\cdot\nabla_{\hat{u}}P(\hat{\mathbf{u}},t) + \tfrac{1}{2}\,(\hat{\mathbf{u}}'-\hat{\mathbf{u}})\,(\hat{\mathbf{u}}'-\hat{\mathbf{u}}) : \nabla_{\hat{u}}\nabla_{\hat{u}}P(\hat{\mathbf{u}},t)\;,$$

where $\nabla_{\hat{u}}$ is the gradient operator in eq.(6), where differentiations are now taken with respect to the Cartesian coordinates $\hat{u}_x$, $\hat{u}_y$, and $\hat{u}_z$ of $\hat{\mathbf{u}}$. Contrary to the similar expansion for translational motion in eq.(5), there is now the restriction that both $\hat{\mathbf{u}}$ and $\hat{\mathbf{u}}'$ are unit vectors. To account for this, note that $\hat{\mathbf{u}}'$ is obtained from $\hat{\mathbf{u}}$ by a rotation, so that $\hat{\mathbf{u}}'-\hat{\mathbf{u}} = \mathbf{w}\times\hat{\mathbf{u}}$, where $\mathbf{w}$ is a vector that is perpendicular to both $\hat{\mathbf{u}}$ and $\hat{\mathbf{u}}'$, which is referred to as *the rotation vector* (see Fig.3b). [4]. Using the identity $(\mathbf{w}\times\hat{\mathbf{u}})\cdot\nabla_{\hat{u}}P = \mathbf{w}\cdot(\hat{\mathbf{u}}\times\nabla_{\hat{u}}P)$, the above Taylor expansion can thus be written as,

$$P(\hat{\mathbf{u}}',t) \;=\; P(\hat{\mathbf{u}},t) + \mathbf{w}\cdot\hat{\mathcal{R}}P(\hat{\mathbf{u}},t) + \tfrac{1}{2}\,\mathbf{w}\mathbf{w} : \hat{\mathcal{R}}\hat{\mathcal{R}}P(\hat{\mathbf{u}},t)\;,$$

where *the rotational operator* is defined as,

$$\hat{\mathcal{R}}\,(\cdots) \;=\; \hat{\mathbf{u}}\times\nabla_{\hat{u}}\,(\cdots)\;. \tag{18}$$

Substitution of this Taylor expansion into the equation of motion leads to,

$$\frac{\partial P(\hat{\mathbf{u}},t)}{\partial\, t} \;=\; \Big[\,\hat{\mathcal{R}}P(\hat{\mathbf{u}},t)\,\Big]\cdot\oint d\hat{u}'\;\mathbf{w}\,W(\hat{\mathbf{u}}'\,|\,\hat{\mathbf{u}}) + \tfrac{1}{2}\,\Big[\,\hat{\mathcal{R}}\hat{\mathcal{R}}P(\hat{\mathbf{u}},t)\,\Big] : \oint d\hat{u}'\;\mathbf{w}\,\mathbf{w}\,W(\hat{\mathbf{u}}'\,|\,\hat{\mathbf{u}})\;.$$

Note that integration with respect to $\hat{\mathbf{u}}'$ is identical to integration with respect to $\mathbf{w}$, where the integration is with respect to the magnitude of $\mathbf{w}$ and with respect to directions of $\mathbf{w}$ within the plane perpendicular to $\hat{\mathbf{u}}$. The first integral vanishes since the transition probability is the same for two oppositely oriented rotations $\mathbf{w}$. An infinitesimal change dw in the length of the rotation vector gives rise to a new unit vector $\hat{\mathbf{u}}$" that is a distance dw away from the original $\mathbf{u}'$ (see Fig.3c), while it is easily seen that a rotation of $\mathbf{w}$ over an angle α changes the direction of $\hat{\mathbf{u}}'$ over the same angle. Hence, $d\hat{u}' = dw\,d\alpha$. Since $W(\hat{\mathbf{u}}'\,|\,\hat{\mathbf{u}})$ is invariant under a rotation of $\mathbf{w}$, and only depends on the magnitude w of the rotation vector, we have (with $W(\hat{\mathbf{u}}'\,|\,\hat{\mathbf{u}}) \equiv W(w)$),

$$\oint d\hat{u}'\;\mathbf{w}\,\mathbf{w}\,W(\hat{\mathbf{u}}'\,|\,\hat{\mathbf{u}}) \;=\; \int_0^{2\pi} d\alpha\;\hat{\mathbf{w}}\,\hat{\mathbf{w}}\times\int_0^{\infty} dw\;w^2\,W(w) \;=\; \pi\,\hat{\mathbf{I}}(\hat{\mathbf{u}})\times\int_0^{\infty} dw\;w^2\,W(w),$$

where $\hat{\mathbf{w}} = \mathbf{w}/w$ is the unit-vector along the rotational vector, and where $\hat{\mathbf{I}}(\hat{\mathbf{u}})$ is the two-dimensional identity tensor within the plane perpendicular to $\hat{\mathbf{u}}$. Since $\hat{\mathcal{R}}P(\hat{\mathbf{u}},t)$ lies in that same plane, we thus finally find *the rotational diffusion equation* (with $\hat{\mathcal{R}}^2(\cdots) = \hat{\mathcal{R}}\cdot\hat{\mathcal{R}}(\cdots)$),

$$\frac{\partial P(\hat{\mathbf{u}},t)}{\partial\, t} \;=\; D_r\,\hat{\mathcal{R}}^2 P(\hat{\mathbf{u}},t)\;, \tag{19}$$

[4] Remember that the outer product $\mathbf{a}\times\mathbf{b}$ of two vectors $\mathbf{a}$ and $\mathbf{b}$ is a vector of length $|\,\mathbf{a}\,|\,|\,\mathbf{b}\,|\;\sin\Theta$, where Θ is the angle between the two vectors, and which has a direction given by the cork-screw rule: "the propagation direction of a cork-screw when it is rotated from $\mathbf{a}$ towards $\mathbf{b}$")

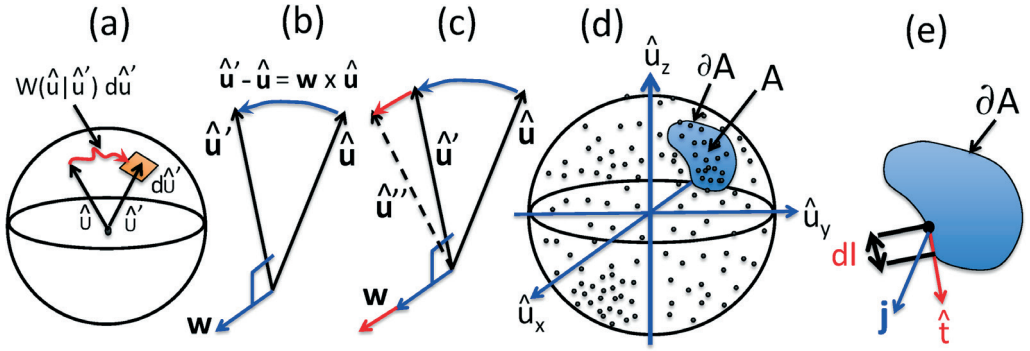


Fig. 3: *(a) The tip of the unit vector* $\hat{\mathbf{u}}$ *has a probability, per unit time, to jump to a surface area element* $d\hat{u}'$ *centered at* $\hat{\mathbf{u}}'$*. (b) For two unit vectors* $\hat{\mathbf{u}}$ *and* $\hat{\mathbf{u}}'$ *that are close together, the difference* $\hat{\mathbf{u}}' - \hat{\mathbf{u}}$ *can be written as* $\mathbf{w} \times \hat{\mathbf{u}}$*, where the rotation vector* $\mathbf{w}$ *is perpendicular to the plane spanned by the two unit vectors. (c) A small increase in the magnitude of the rotation vector (as indicated in red) leads to an increase of the distance between* $\hat{\mathbf{u}}$ *and the new unit vector* $\hat{\mathbf{u}}''$ *(also indicated in red). (d) An arbitrary surface area* A *on the unit spherical surface, with boundary* ∂A*. (d) The unit vector* $\hat{\mathbf{t}}$ *is locally tangential to the curve* ∂A*, while* dl *is an infinitesimally small part of* ∂A*.*

where,

$$D_r \;=\; \frac{\pi}{2}\int_0^\infty dw\; w^2\, W(w)\;, \tag{20}$$

is referred to as *the rotational diffusion coefficient*. The diffusion coefficient in Fick's law (7) is sometimes referred to as *the translational diffusion coefficient*.
The time dependence of the average of the orientation vector $\hat{\mathbf{u}}$ can be obtained from eq.(19) by multiplying both sides with $\hat{\mathbf{u}}$ and integration over the unit-spherical surface,

$$\frac{d<\hat{\mathbf{u}}>(t)}{dt} \;=\; D_r \oint d\hat{\mathbf{u}}\,\hat{\mathbf{u}}\,\hat{\mathcal{R}}^2 P(\hat{\mathbf{u}},t) \;=\; D_r \oint d\hat{\mathbf{u}}\,P(\hat{\mathbf{u}},t)\,\hat{\mathcal{R}}^2\hat{\mathbf{u}} \;=\; -2\,D_r\;<\hat{\mathbf{u}}>(t),$$

where a partial integration is performed in the second equation (for mathematical details, see appendix B), and where $\hat{\mathcal{R}}^2\hat{\mathbf{u}} = -2\hat{\mathbf{u}}$ is used to obtain the last equation. Let $\hat{\mathbf{u}}_0$ denote the unit vector at time zero. The orientational equivalent of mean-squared displacement as introduced in subsection 2.1 for translational motion is $W_{\hat{u}}(t) = <|\,\hat{\mathbf{u}} - \hat{\mathbf{u}}_0\,|^2>$. From the above result it is found that,

$$W_{\hat{u}}(t) \;=\; 2\,D_r\,\left[1 - \exp\left\{-2\,D_r\,t\right\}\right]\;.$$

For small times. such that $D_r t \ll 1$, $W_{\hat{u}}(t)$ is the displacement of the tip of $\hat{\mathbf{u}}$ on the two-dimensional plane that is perpendicular to $\hat{\mathbf{u}}_0$. Expanding the exponent (using that $\exp x = 1 + x + \cdots$) leads to $W_{\hat{u}}(t) = 4\,D_r t$, which indeed complies to eq.(9) for translational diffusion in two dimensions.

3.2 The orientational continuity equation

Similar to what has been done in section 2.2 for translational diffusion, we derive a continuity equation by counting the number of points that leave/enter an arbitrary surface area A, through

the closed curve ∂A that bounds the area A (see Fig.3d). The rate of change of the "number of points" $N(t)$ within the surface area A is equal to,

$$\frac{dN(t)}{dt} \;=\; \int_A d\hat{u}\; \frac{\partial}{\partial t}\, P(\hat{\mathbf{u}},t)\;, \tag{21}$$

where, as already mentioned above, $d\hat{u}$ denotes a small surface area on the unit-spherical surface. Let dl denote an infinitesimally small arclength on the curve ∂A, and let $\hat{\mathbf{t}}$ denote the unit vector that is tangential to the curve ∂A in anti-clockwise orientation (as depicted in Fig.3e). The flux of points due to rotation of the particles is equal to $\mathbf{j} = P(\hat{\mathbf{u}},t)\,\partial\hat{\mathbf{u}}/\partial t$. The component of this flux that gives rise to changes of the number of points in A is the component perpendicular to $\hat{\mathbf{t}}$, and of course locally parallel to the unit spherical surface. The unit vector in that direction is equal to $\hat{\mathbf{u}}\times\hat{\mathbf{t}}$. The total number of points that pass through ∂A per unit time is thus equal to (with $d\mathbf{l} \;=\; \hat{\mathbf{t}}\,dl$),

$$\frac{dN(t)}{dt} \;=\; \oint_{\partial A} dl\;(\hat{\mathbf{u}}\times\hat{\mathbf{t}})\cdot\left[\,P(\hat{\mathbf{u}},t)\,\frac{\partial\hat{\mathbf{u}}}{\partial t}\,\right] \;=\; -\oint_{\partial A} d\mathbf{l}\cdot\left[\,P(\hat{\mathbf{u}},t)\;\hat{\mathbf{u}}\times\frac{\partial\hat{\mathbf{u}}}{\partial t}\,\right]\;, \tag{22}$$

where it is used that $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})$, and $\mathbf{c}\times\mathbf{a} = -\,\mathbf{a}\times\mathbf{c}$ (for any vectors, $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$). The integral in the above equation is a line integral over the closed curve ∂A. The integral over the closed curve ∂A can be written in terms of a surface integral ranging over its interior A, using Stokes's integral theorem (see appendix B). Applying Stokes's theorem to eq.(22), and comparing with eq.(21) leads to,

$$\int_A d\hat{u}\;\left[\,\frac{\partial}{\partial t}\,P(\hat{\mathbf{u}},t) + \hat{\mathbf{u}}\cdot\left\{\nabla\times\left(\,P(\hat{\mathbf{u}},t)\;\hat{\mathbf{u}}\times\frac{\partial\hat{\mathbf{u}}}{\partial t}\,\right)\right\}\right] \;=\; 0\;.$$

Since this holds for arbitrary surfaces A, the integrand must be zero. Using the identity $\hat{\mathbf{u}}\cdot\nabla_{\hat{u}}\times(\cdots) = (\hat{\mathbf{u}}\times\nabla_{\hat{u}})\cdot(\cdots)$, we thus find that,

$$\frac{\partial}{\partial t}\,P(\hat{\mathbf{u}},t) \;=\; -\,\hat{\mathcal{R}}\cdot[\,\mathbf{\Omega}(\hat{\mathbf{u}},t)\,P(\hat{\mathbf{u}},t)\,]\;, \tag{23}$$

where the *orientational operator* is defined in eq.(18), while the (thermally averaged) *rotational velocity* is equal to,

$$\mathbf{\Omega} \;=\; \hat{\mathbf{u}}\times\frac{\partial\,\hat{\mathbf{u}}}{\partial t}\;. \tag{24}$$

The orientational velocity is defined through $\partial\hat{\mathbf{u}}/\partial\,t = \mathbf{\Omega}\times\hat{\mathbf{u}}$. Taking the outer product of both sides with $\hat{\mathbf{u}}$, using the identity $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}$, and noting that $\mathbf{\Omega}\perp\hat{\mathbf{u}}$ (so that $\mathbf{\Omega}\cdot\hat{\mathbf{u}} = 0$), eq.(24) is recovered. The result in eq.(23) is the orientational analogue of the continuity equation (13) for translational motion.

3.3 The rotational diffusion coefficient, and the Brownian torque

The hydrodynamic torque $\mathbf{T}^h$ (where the superscript "h" stands for "hydrodynamic") resulting from friction with the fluid, for a "sufficiently symmetric" colloidal particle, is proportional and opposite in direction to the rotational velocity $\mathbf{\Omega}$. This can be seen as follows. The general definition of a torque is $\mathbf{T} \;=\; \oint dS\,\mathbf{r}\times\mathbf{F}$, where the integral ranges over the surface of the

colloid, and where $\mathbf{F}$ is the force per unit area acting on a local surface element at position $\mathbf{r}$. For a rotating colloidal particle, the friction force $\mathbf{F}$ that the fluid exerts onto surface elements is opposite to the velocity of that surface element: $\mathbf{F} \sim -\mathbf{v} = -\mathbf{\Omega} \times \mathbf{r}$. Hence, $\mathbf{T} \sim -\oint dS\, \mathbf{r} \times (\mathbf{\Omega} \times \mathbf{r}) = -\oint dS\, [r^2\, \mathbf{\Omega} - (\mathbf{\Omega} \cdot \mathbf{r})\, \mathbf{r}]$. For a spherical particle we have $\oint dS\, \mathbf{rr} = (4\pi/3) r^2 \hat{\mathbf{I}}$. A similar relation holds for platelets (where $\mathbf{\Omega}$ lies within the plane spanned by the face of the platelet), and for uniaxial rod-like particles (where $\mathbf{\Omega}$ is perpendicular to $\hat{\mathbf{u}}$). It thus follows that $\mathbf{T} \sim -\mathbf{\Omega}$, with a positive proportionality constant. Hence, $\mathbf{T}^h = -\zeta_r\, \mathbf{\Omega}$, where the positive proportionality constant ζ_r is referred to as *the rotational friction coefficient*. From torque balance on the diffusive time scale we know that the total torque is zero: $\mathbf{T}^{tot} = \mathbf{T} - \zeta_r\, \mathbf{\Omega} = \mathbf{0}$, with $\mathbf{T}$ represents the non-hydrodynamic torques, and hence $\mathbf{\Omega} = \mathbf{T}/\zeta_r$. According to the rotational continuity equation (23) it thus follows that,

$$\frac{\partial}{\partial t}\, P(\hat{\mathbf{u}}, t) \;=\; -\frac{1}{\zeta_r}\, \hat{\mathcal{R}} \cdot [\, P(\hat{\mathbf{u}}, t)\, \mathbf{T}(\hat{\mathbf{u}}, t)\,] \;.$$

Comparing with Fick's rotational diffusion equation (19), it is found that the torque in the absence of an external field is equal to $\mathbf{T} = -(\zeta_r D_r)\, \hat{\mathcal{R}} \ln P(\hat{\mathbf{u}}, t)$, which torque will be further addressed at the end of this subsection. In the presence of an additional torque $\mathbf{T}^{ext}$ due to an external field, the torque in the above equation is equal to $\mathbf{T} = -(\zeta_r D_r)\, \hat{\mathcal{R}} \ln P(\hat{\mathbf{u}}, t) + \mathbf{T}^{ext}$. The external field is described by a position and orientation dependent potential Ψ. Whereas the force is equal to $-\nabla\Psi$, the torque is equal to $-\hat{\mathcal{R}}\Psi$. The above equation of motion thus becomes,

$$\frac{\partial}{\partial t}\, P(\hat{\mathbf{u}}, t) \;=\; \hat{\mathcal{R}} \cdot \left[\, D_r\, \hat{\mathcal{R}} P(\hat{\mathbf{u}}, t) + \frac{1}{\zeta_r}\, P(\hat{\mathbf{u}}, t)\, \hat{\mathcal{R}} \Psi(\hat{\mathbf{u}}) \right] \;.$$

In equilibrium, where $\partial P/\partial\, t = 0$, the combination between the square brackets is zero, and hence $P \sim \exp\{-\Psi/(\zeta_r\, D_r)\}$. From equilibrium statistical mechanics we know that this should be equal to $\sim \exp\{-\Psi/k_B T\}$, from which it follows that,

$$D_r \;=\; \frac{k_B T}{\zeta_r} \;, \tag{25}$$

which is the rotational counter part of Einstein's expression for the translational diffusion coefficient in eq.(16). The above mentioned force in the absence of an external field is thus equal to,

$$\mathbf{T}^{Br} \;=\; -k_B T\, \hat{\mathcal{R}} \ln P(\hat{\mathbf{u}}, t) \;. \tag{26}$$

which will be referred to as *the Brownian torque*. Like the Brownian force in eq.(17), this torque is of an entropic origin.

4 Applications: Combined Translational and Rotational Diffusion in the Presence of an Electric Field, Active Motion, and Shear Flow

In this section, three applications of the above derived diffusion equations will be discussed: polarization of a dispersion of rod-like colloids in an oscillating electric field, diffusion of an active swimmer, and of Taylor-dispersion of spherical particles in simple shear flow.

The appropriate diffusion equation is generally an equation of motion for combined translational and rotational motion for the pdf $P(\mathbf{r},\hat{\mathbf{u}},t)$. This pdf is related in an obvious way to the probability for the center-of-mass $\mathbf{r}$ and the orientation $\hat{\mathbf{u}}$ of a colloidal particle at time t. The average of any function $f(\mathbf{r},\hat{\mathbf{u}})$ is now given by,

$$< f(\mathbf{r},\hat{\mathbf{u}}) > (t) \;=\; \int d\mathbf{r} \oint d\hat{\mathbf{u}}\; f(\mathbf{r},\hat{\mathbf{u}})\, P(\mathbf{r},\hat{\mathbf{u}},t)\;,$$

where, as before, the integrations range over the entire three-dimensional space and the unit-spherical surface, respectively.
The equation of motion for this pdf is simply given by the sum of the rate-of-change due to translational and rotational motion. According to the continuity equations (13) and (23), we thus have,

$$\frac{\partial}{\partial\, t}\, P(\mathbf{r},\hat{\mathbf{u}},t) \;\;=\;\; -\,\nabla\cdot[\,\mathbf{v}\,P(\mathbf{r},\hat{\mathbf{u}},t)\,] - \,\hat{\mathcal{R}}\cdot[\,\mathbf{\Omega}\,P(\mathbf{r},\hat{\mathbf{u}},t)\,]\;, \tag{27}$$

where the translational velocity $\mathbf{v}$ and the rotational velocity $\mathbf{\Omega}$ can be obtained from force- and torque balance, together with the expressions that we derived above for the Brownian force and torque. This features quite simple and straightforward derivations of diffusion equations in the presence of external fields, for active swimmers, and for shear flow, to mention a few.

4.1 Electric permittivity

Molecules that carry a permanent electric dipole moment will be aligned by an external electric field $\mathbf{E}$. The definition of the electric dipole moment of a single particle is,

$$\mathbf{p} \;=\; \int_{V_p} d\mathbf{r}\; \rho(\mathbf{r})\,\mathbf{r}\;, \tag{28}$$

where the integral ranges over the volume V_p of the particle, and where $\rho(\mathbf{r})$ is the charge density within the particle (that is, the charge per unit volume). Since the force on a charge q in an electric field $\mathbf{E}$ is equal to $q\,\mathbf{E}$, so that the force per unit volume is equal to $\mathbf{F}=\rho(\mathbf{r})\,\mathbf{E}$, the torque on the dipole is equal to,

$$\mathbf{T} \;=\; \int_{V_p} d\mathbf{r}\;\mathbf{r}\times\mathbf{F} \;=\; \int_{V_p} d\mathbf{r}\;\mathbf{r}\times[\,\rho(\mathbf{r})\,\mathbf{E}\,] \;=\; \left[\int_{V_p} d\mathbf{r}\;\rho(\mathbf{r})\,\mathbf{r}\right]\times\mathbf{E} \;=\; \mathbf{p}\times\mathbf{E}\;.$$

As an assembly of N particles is subjected to an external electric field, this torque gives rise to alignment of the particles. This induces a macroscopic dipole density $\mathbf{P}$, that is, the total dipole moment per unit volume. The local dipole density is simply the sum of all dipoles of the particles that are inside a (small) volume V, divided by that volume,

$$\mathbf{P} \;=\; \frac{1}{V}\sum_{i=1}^{N} <\mathbf{p}_i> = \;\frac{N}{V}\;<\mathbf{p}>\;\;.$$

In the last equation, identical particles are assumed, so that $<\mathbf{p}_i>$ is the same for all particles and is denoted as $<\mathbf{p}>$. For sufficiently small electric field amplitudes the induced dipole density is proportional to the applied field strength,

$$\mathbf{P} \;=\; \chi\,\mathbf{E}\,, \tag{29}$$

where χ is referred to as *the electric permittivity* (which in turn is related to the dielectric constant). In this subsection the interest is in the evaluation of the permittivity for oscillating electric fields.
The equation of motion for the electric dipole is obtained as follows. First of all, to describe rotational motion, we need only to consider the rotational diffusion equation (23), instead of the full combined translational-rotational equation of motion (27). According to the discussion in section 1.2, together with the identification of the Brownian torque in subsection 3.3, the torque-balance equation is, $-\zeta_r\,\mathbf{\Omega} - k_BT\nabla \ln P + \mathbf{T} = \mathbf{0}$, and hence (with $\beta = 1/k_BT$),

$$\mathbf{\Omega} \;=\; \beta\, D_r \,\Big[-k_BT\,\hat{\mathcal{R}} \ln P(\mathbf{r},\hat{\mathbf{u}},t) + \mathbf{T}\Big] \;.$$

Substitution of this expression for the rotational velocity into eq.(23) leads to,

$$\frac{\partial}{\partial\, t}\, P(\hat{\mathbf{u}},t) \;\;=\;\; D_r\hat{\mathcal{R}}\cdot\Big[\hat{\mathcal{R}}P(\hat{\mathbf{u}},t) - \beta\,\mathbf{T}\,P(\hat{\mathbf{u}},t)\Big] \;. \tag{30}$$

The orientation of the dipole moment is $\hat{\mathbf{u}}$, so that we can write,

$$\mathbf{p} \;=\; p\,\hat{\mathbf{u}}\;,$$

where p is the magnitude of the dipole moment. We consider an oscillating electric field,

$$\mathbf{E};= \;\mathbf{E}_0\;\cos\{\omega\, t\}\;,$$

with $\mathbf{E}_0$ the constant field amplitude, and ω the frequency (which is equal to $2\,\pi\,\nu$, where $\nu = 1/\tau$, with τ the time for a single cycle of the electric field). To within linear response, the pdf is written as (with $E_0 =\mid \mathbf{E}_0\mid$),

$$P(\hat{\mathbf{u}},t) \;=\; P_0(\hat{\mathbf{u}}) + E_0\,P_1(\hat{\mathbf{u}},t)\;, \tag{31}$$

where P_0 is the pdf in the absence of the electric field, and P_1 the electric-field induced perturbation of the pdf. The pdf P_0 is a constant,

$$P_0(\hat{\mathbf{u}}) \;=\; \frac{1}{4\,\pi}\;,$$

which follows from normalization. An equation of motion for $<\hat{\mathbf{u}}>\,(t)$, and hence for the electric dipole moment, is obtained from the equation of motion (30) by multiplying both sides with $\hat{\mathbf{u}}$ and integration over all orientations,

$$\begin{aligned}\frac{d<\hat{\mathbf{u}}>(t)}{d\,t} \;&=\; D_r\oint d\hat{\mathbf{u}}\;\;\hat{\mathbf{u}}\,\hat{\mathcal{R}}\cdot\Big[\hat{\mathcal{R}}P(\hat{\mathbf{u}},t) - \beta\,\mathbf{T}\,P(\hat{\mathbf{u}},t)\Big]\\ &=\; D_r\oint d\hat{\mathbf{u}}\;P(\hat{\mathbf{u}},t)\;\Big[\hat{\mathcal{R}}^2\,\hat{\mathbf{u}} + \beta\,p\,(\hat{\mathbf{u}}\times\mathbf{E})\cdot\hat{\mathcal{R}}\,\hat{\mathbf{u}}\Big]\;,\end{aligned}$$

where in the second line partial integrations have been performed (as explained in appendix B). Using the identities $\hat{\mathcal{R}}^2\,\hat{\mathbf{u}} = -2\,\hat{\mathbf{u}}$ and $(\hat{\mathbf{u}}\times\mathbf{E})\cdot\hat{\mathcal{R}}\,\hat{\mathbf{u}} = (\hat{\mathbf{I}} - \hat{\mathbf{u}}\hat{\mathbf{u}})\cdot\mathbf{E}$, we thus have,

$$\frac{d<\hat{\mathbf{u}}>(t)}{d\,t} \;\;=\;\; -\,D_r\Big[\,2<\hat{\mathbf{u}}>(t) - \beta\,p\,\Big(\hat{\mathbf{I}} - <\hat{\mathbf{u}}\hat{\mathbf{u}}>(t)\Big)\cdot\mathbf{E}_0\,\cos\{\omega\,t\}\Big]\;.$$

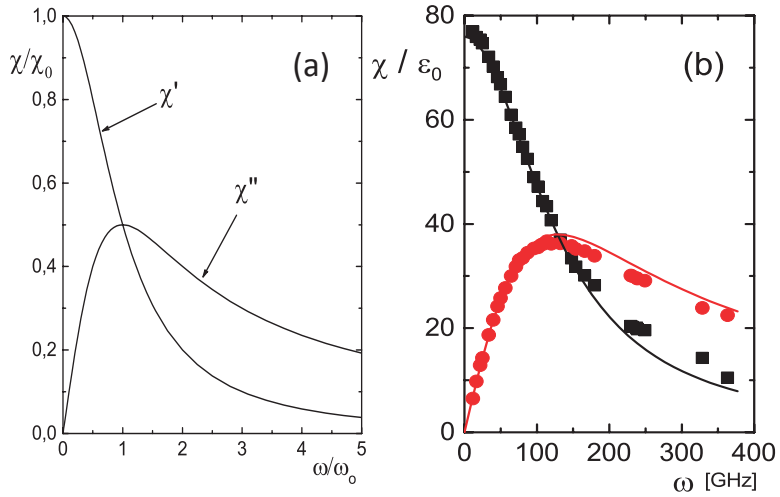


Fig. 4: *(a) The in-phase and out-phase response functions in eq.(32) as a function of the frequency. (b) The response functions for pure water. Data are taken from Ref.[6].*

To linear order in the electric field amplitude, the average $<\hat{\mathbf{u}}\hat{\mathbf{u}}>(t)$ can be evaluated with respect to the pdf P_0 in the expansion (31), which gives $<\hat{\mathbf{u}}\hat{\mathbf{u}}>(t) = (1/3)\,\hat{\mathbf{I}}$. Furthermore, due to symmetry, only the component of $\hat{\mathbf{u}}$ along the electric field will be non-zero on average. Taking the direction of the electric field along the z-direction thus leads to the following equation of motion for the z-component of $\hat{\mathbf{u}}$,

$$\frac{d<\hat{u}_z>(t)}{dt} \;=\; -D_r\left[\,2<\hat{u}_z>(t) - \tfrac{2}{3}\beta p E_0\,\cos\{\omega t\}\,\right] .$$

while $<\hat{u}_x>=0=<\hat{u}_y>$. The solution of this equation is,

$$<\hat{u}_z>(t) \;=\; \frac{2}{3}\,\frac{\beta p D_r}{(2D_r)^2+\omega^2}\,\left[\,2D_r\,\cos\{\omega t\} + \omega\,\sin\{\omega t\}\,\right] E_0 .$$

There is thus an in-phase component $\sim \cos\{\omega t\}$ to the electric field and an out-phase component $\sim \sin\{\omega t\}$. For frequencies larger than $2D_r$ an out-phase component develops as the orientation of rods can not follow the field instantaneously. It is therefore necessary to distinguish between the in-phase and out-phase response, and to reformulate eq.(29) as,

$$\mathbf{P}(\omega,t) \;=\; \left[\,\chi'(\omega)\,\cos\{\omega t\} + \chi''(\omega)\,\sin\{\omega t\}\,\right]\mathbf{E}_0 ,$$

where the in-phase χ' and out-phase χ'' response functions are equal to,

$$\begin{aligned}\chi'(\omega) &= \chi_0\,\frac{\omega_0^2}{\omega_0^2+\omega^2} ,\\ \chi''(\omega) &= \chi_0\,\frac{\omega_0\,\omega}{\omega_0^2+\omega^2} ,\end{aligned} \qquad (32)$$

where the response amplitude χ_0 and the characteristic frequency ω_0 are equal to,

$$\chi_0 \;=\; \frac{N}{V}\,\frac{1}{3}\,\beta p^2 \quad , \quad \omega_0 \;=\; 2D_r . \qquad (33)$$

The frequency dependence of response functions as found in eq.(32) are referred to as *Debye-Maxwell response functions*. The two response functions are plotted in Fig.4. For low frequencies there is essentially instantaneous response, so that χ' is constant, while χ'' is zero. At frequencies of the order of the characteristic frequency ω_0, the dynamics of the dipoles is too slow to instantaneously adapt to the external field, so that a significant out-phase component comes into play, while the in-phase response function decreases. At very high frequencies the dipoles can not respond anymore, so that both response functions become zero. This is a typical scenario for many types of response functions to various external oscillating fields.
Experimental data for the electric permittivity of pure water at $25^o\,C$ are given in Fig.4b (where the permittivity is dimensionalized with the dielectric constant ϵ_0 of vacuum). These data are taken from Ref.[6]. The solid lines correspond to eq.(32) with $\chi_0/\epsilon_0 = 76$ and $\omega_0 = 128\,GHz = 128\times 10^9\,Hz$. The functional dependence on frequency quite satisfactory described by the above theory, with some deviations at high frequencies. Note that the geometry of a water molecule is not rod-like, as assumed in the above analysis. As the two rotational diffusion coefficients related to the two axis of rotation that give rise to a change of the orientation of the dipole are not too different, the theory remains valid on a semi-quantitative level. What is not accounted for in the above analysis are interactions between the molecules, which certainly play an important role in water. The dynamics of a water molecule is affected by interactions, and the field experienced by each molecule is generally different from the applied field. The field that is experienced by individual molecules, the so-called local electric field, is the sum of the externally applied field plus the electric fields generated by the surrounding molecules. These features render an analysis at higher concentrations much more involved, and is beyond the scope of this lecture. The rotational dynamics of water molecules proceeds on the pico-second ($= 10^{-12}\,s$) time scale, which implies that $\omega_0 \approx 1000 \times 10^9\,Hz$, while the dipole moment of water is $6\times 10^{-30}\,C\,m$, which implies that $\chi_0/\epsilon_0 = 11$ (using that $\epsilon_0\,k_B\,T = 3.6\times 10^{-32}\,C^2/m$ and $N/V = 3.3\times 10^{28}\,m^{-3}$). These values are of the same order of magnitude as found from the fits. The deviations are most probably due to water-water interactions.

4.2 Diffusion and active motion of a spherical microswimmer

There are different types of colloidal-like entities that are self-propelled by means of forces that they exert onto the surrounding solvent, leading to "active motion". Examples are bacteria with actively moving flagella, sperm, molecular motors, and synthetic microswimmers of which an example will be discussed later.
The equation of motion for an active swimmer can be obtained as follows. The swimming velocity has a constant magnitude v_0, while the swimming direction is given by the unit-vector $\hat{\mathbf{u}}$, which is set by the orientation of the swimmer. For a spherical particle, the orientation is determined through the surface properties that lead to self propulsion.
The discussion below is restricted to spherical swimmers, but the entire analysis can be extended to non-spherical swimmers. The present analysis can also easily be extended to include external (alignment) fields.
Note that the pdf $P(\hat{\mathbf{u}},t)$ is no longer invariant under inversion of the direction of $\hat{\mathbf{u}}$, as it is now connected to the swimming direction: $P(\hat{\mathbf{u}},t) \neq P(-\hat{\mathbf{u}},t)$. The friction force with the fluid is now equal to $\mathbf{F}^h = -\zeta\,[\,\mathbf{v} - v_0\,\hat{\mathbf{u}}\,]$. Force balance (without an external field) implies that $-\zeta\,[\,\mathbf{v} - v_0\,\hat{\mathbf{u}}\,] - k_BT\nabla \ln P = \mathbf{0}$, and hence,

$$\mathbf{v} = v_0\,\hat{\mathbf{u}} - k_BT\,\nabla \ln P(\mathbf{r},\hat{\mathbf{u}},t)\ .$$

For swimmers that would swim in a straight line in the absence of Brownian motion, torque balance is not affected by self propulsion. Circle-swimmers are not considered here. According to eq.(27), the diffusion equation for a spherical, active swimmer thus reads,

$$\frac{\partial}{\partial t}\,P(\mathbf{r},\hat{\mathbf{u}},t) \;=\; -\nabla\cdot[\,v_0\,\hat{\mathbf{u}}\,P(\mathbf{r},\hat{\mathbf{u}},t)\,] + D\,\nabla^2 P(\mathbf{r},\hat{\mathbf{u}},t) + \,D_r\,\hat{\mathcal{R}}^2 P(\mathbf{r},\hat{\mathbf{u}},t)\;. \quad (34)$$

An experimentally accessible quantity is the mean-squared displacement $W(t) =< r^2 > (t)$. Without self propulsion $W(t) = 6\,D\,t$ (see eq.(9)). Swimming will certainly change this result, and will depend on the magnitude v_0 of the swimming velocity and the rate-of-change of orientation of the swimming direction that is characterized by the rotational diffusion coefficient D_r. Multiplying both sides of the equation of motion (34) with r^2, and subsequent (partial) integration leads to (partial integrations are discussed in appendix B),

$$\frac{dW(t)}{d\,t} \;=\; 6\,D + 2\,v_0\int\!d\mathbf{r}\oint\!d\hat{\mathbf{u}}\;(\hat{\mathbf{u}}\cdot\mathbf{r})\,P(\mathbf{r},\hat{\mathbf{u}},t) \;=\; 6\,D + 2\,v_0\;<\hat{\mathbf{u}}\cdot\mathbf{r}>(t)\;.$$

In order to solve this equation of motion, the time dependence of the average on the right hand-side needs be evaluated. Multiplying both sides of eq.(34) with $\hat{\mathbf{u}}\cdot\mathbf{r}$ and integration, performing again a partial integration, and using the identity $\hat{\mathcal{R}}^2\hat{\mathbf{u}} = -2\,\hat{\mathbf{u}}$, gives,

$$\frac{d<\hat{\mathbf{u}}\cdot\mathbf{r}>(t)}{d\,t} \;=\; v_0 - 2\,D_r\;<\hat{\mathbf{u}}\cdot\mathbf{r}>(t)\;.$$

The solution of this equation is (note that at time $t = 0$ the particle is assumed to be at the origin),

$$<\hat{\mathbf{u}}\cdot\mathbf{r}>(t) \;=\; \frac{v_0}{2\,D_r}\,[\,1-\exp\{-2\,D_r\,t\,\}\,]\;.$$

Substitution of this result into the above equation of motion for the mean-squared displacement leads to a closed equation, the solution of which is,

$$W(t) \;=\; 6\left[\,D+\frac{v_0^2}{6\,D_r}\,\right]t + \frac{v_0^2}{2\,D_r^2}\,[\,\exp\{-2\,D_r\,t\}-1\,]\;. \quad (35)$$

It is instructive to consider two limits. For small times, such that $t \ll 1/D_r$, during which time the particle does not change its orientation, the exponent can be expanded up to second order (using that $\exp\{x\} = 1 + x + \frac{1}{2}x^2 + \cdots$), leading to,

$$W(t) \;=\; 6\,D\,t + v_0^2\,t^2 + \cdots \qquad (\,t \ll 1/D_r\,)\;,$$

where the remaining terms are of higher order in time. The first term is just diffusion without swimming, while the second term is the squared displacement due to a constant velocity that does not change its direction. For large times $t \gg 1/D_r$, for which the particle made many rotations, it is found that,

$$W(t) \;=\; 6\,D_{eff}\,t\;\;,\;\;\text{where}\quad D_{eff} = D + \frac{v_0^2}{6\,D_r} \qquad (\,t \gg 1/D_r\,) \quad (36)$$

The particle thus diffuses as a non-active particle, but with and effective, larger diffusion coefficient.

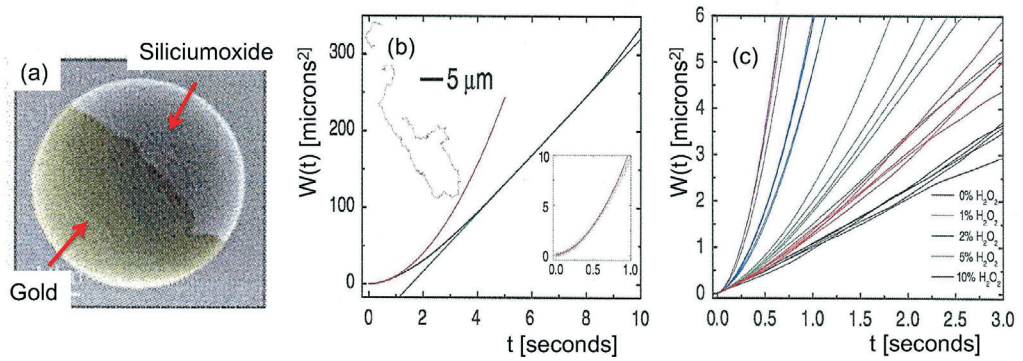


Fig. 5: *(a) An electron microscopy image of a spherical Janus particle with a radius of* $2.1\,\mu m$ *that consists of a core of paramagnetic siliciumoxide, capped on one side by a gold layer with a thickness of* $20\,nm$*. This image is taken from Ref.[9]. (b) The mean-squared displacement of a spherical Janus particle with a radius of* $1.6\,\mu m$ *consisting of polystyrene-latex, coated on one side with a platinum layer with thickness* $5.5\,nm$*. The particle is embedded in a peroxide solution that provides the swimming force through the catalysis of dissociation of the peroxide by platinum. The red line is a fit of the mean-squared displacement to a parabolic time dependence (*$W(t) \sim t^2$*) for short times (see also the insert), while the blue line is a linear fit to the long-time behaviour. The insert on the top left is a typical trajectory. (c) The mean-squared displacement for several peroxide concentrations. The plots in (b) and (c) are taken from Ref.[8]*

This type of diffusive motion is discussed in, for example, Refs.[7, 8, 9, 10]. Experiments on the mean-square displacement have been reported on spherical Janus particles, where one half of the sphere is covered by either gold or platinum [8, 9, 10]. Gold-capped particles that are embedded in a fluid mixture that is close to the gas-liquid critical point are self-propelled due to local demixing as a result of laser-light heating of the gold cap [9, 10]. Figure 5a shows an electron microscopy image of a gold-capped Janus particle (which is taken from Ref.[9]). The intensity of the light source, as well as the size of the spheres, determine the magnitude of the swimming velocity. Alternatively, platinum-coated Janus particles become self-propelling when embedded in a peroxide (H_2O_2) solution in water. Platinum catalysis the decomposition of peroxide in water and oxygen, which leads to an enhancement of the local solvent osmotic pressure on the platinum side, that drives the motion of the Janus particle. Figure 5b shows a fit of the mean-squared displacement to eq.(35), together with it's limiting form for short and long times. Figure 5c shows the mean-squared displacements for several peroxide concentrations. Fits of these curves to eq.(35) reveal that v_0 is linearly dependent on the peroxide concentration for sufficiently low concentrations, that D is independent of the concentration, as expected, while D_r significantly increases with peroxide concentration. The latter observation is probably due to inhomogeneous platinum coverage (similar to what is seen in Fig.5a for gold-capped spheres), which induces an additional rotational motion.

An overview of the current status in the research field of microswimmers, which contains a description of biological and synthetic swimmers, a detailed description of various theoretical models, swimming near surfaces, hydrodynamic synchronization of flagella, as well as collective phenomena in suspensions of swimmers, can be found in Ref.[11].

4.3 Diffusion in shear flow: Taylor dispersion

Here we analyze the diffusion of spherical particles in shear flow, and address what is referred to as *Taylor-dispersion* (first described by the British scientist Geoffrey Taylor). We consider simple shear flow, which allows for a quite simple and full analytical treatment without any mathematical approximations. A simple shear flow is a flow that is generated by displacing two parallel plates relative to each other with a constant velocity. Without loss of generality we chose the two plates to span the xz-plane, and the velocities of the upper and lower plate in the positive and negative x-direction, respectively. The resulting flow velocity of a fluid in between the plates is in the x-direction (*the flow-direction*), and increases linearly in the y-direction (*the gradient-direction*), as depicted by the red arrows in the left panel in Fig.6. The z-direction (or, *the vorticity-direction*) is perpendicular to the sketch in Fig.6. Consider a spherical volume (which we will refer to as a "droplet"). The initial condition for the probability density function of the center-of-mass position of the spherical colloidal particle is taken constant within that volume, and zero outside.

Note that the concentration of colloids is proportional to the pdf of the position coordinate of single colloidal particle, so that one can think of the equation of motion for the pdf as an equation for the concentration, and the droplet as consisting of a dilute suspension embedded in a pure fluid.

Without shear flow, the size of the droplet (that is, the region where the pdf is significantly non-zero) increases due to diffusion as described in subsection 2.1 (see the lower right panel in Fig.6). With flow, colloidal particles are dragged along with the fluid, so that the originally spherical volume will be deformed by convection (see the upper right panel in Fig.6). There are now spatial gradients along a larger circumference as compared to the no-flow case. This leads to additional diffusive mass transport away from the deformed droplet, and thus to an increased "homogenization" (="dispersion"). This flow-induced increased dispersion is referred to as *Taylor dispersion*.

In order to describe Taylor-dispersion in more detail, we need to derive the equation of motion for the pdf for the center-of-mass position of a colloidal particle. Such an equation of motion is referred to as an *advection-diffusion equation*, as both flow and diffusion lead to mass transport. Let $\mathbf{u}(\mathbf{r})$ denote the externally imposed fluid flow velocity. Since a colloidal particle is dragged along by the flowing fluid, there is now an additional translational velocity equal to $\mathbf{u}(\mathbf{r})$ at the position $\mathbf{r}$ of the colloidal particle. The friction force is now equal to $\mathbf{F}^h = -\zeta\,(\mathbf{v} - \mathbf{u}(\mathbf{r}))$, so that force balance gives, $-\zeta\,(\mathbf{v} - \mathbf{u}(\mathbf{r})) - k_BT\nabla \ln P = \mathbf{0}$. Hence,

$$\mathbf{v} \;=\; \mathbf{u}(\mathbf{r}) - k_BT\nabla \ln P(\mathbf{r},t)\;.$$

From eq.(27) for translational motion of a spherical particle, the following advection-diffusion equation is obtained,

$$\frac{\partial}{\partial t}\,P(\mathbf{r},t) \;\;=\;\; -\nabla\cdot[\,\mathbf{u}(\mathbf{r})\,P(\mathbf{r},t)\,] + D\,\nabla^2 P(\mathbf{r},t)\;.$$

For the simple shear flow as discussed above (see the left panel in Fig.6), the fluid flow velocity is equal to $\mathbf{u}(\mathbf{r}) = \dot\gamma\, y\,\hat{\mathbf{e}}_x$, where *the shear rate* $\dot\gamma = v_{plate}/L$ quantifies the spatial gradient of the flow velocity (with v_{plate} the velocity of the upper plate, and $2L$ the distance between the plates), and where $\hat{\mathbf{e}}_x$ is the unit vector along the x-direction. The above advection-diffusion equation for such a flow reduces to,

$$\frac{\partial}{\partial t}\,P(\mathbf{r},t) \;\;=\;\; -\dot\gamma\, y\,\frac{\partial}{\partial x}P(\mathbf{r},t) + D\,\nabla^2 P(\mathbf{r},t)\;. \qquad (37)$$

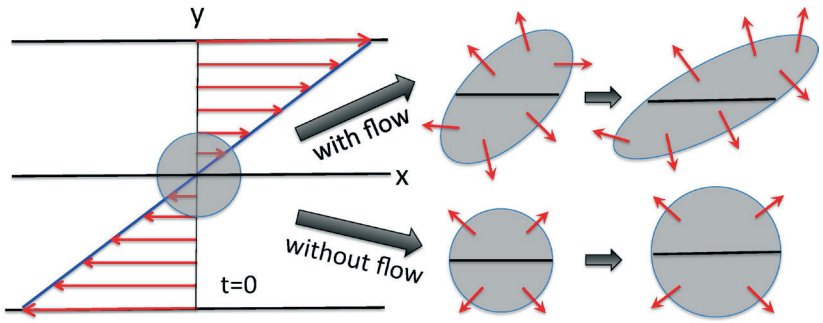


Fig. 6: *Shear flow is induced by moving parallel plates relative to each other. At time zero, say, the pdf has a constant, non-zero value within the grey spherical region at the origin, and is zero outside that region. Without flow the size of sphere where the pdf is significantly non-zero increases due to diffusion (as illustrated in the lower panel). With shear flow the spherical region is convected to a cigar-like object (upper panel), where the flow-induced increase of the circumference leads to additional diffusion (indicated by the red arrows), which leads to an increased "homogenization", or "dispersion".*

For convenience we will assume the initial size of the droplet to be mathematically infinitely small, that is, the particle is assumed to be located at the origin at time zero. Furthermore, we analyze flow-enhanced dispersion for droplet sizes which are much smaller than the distance between the plates, so that the droplet is in a quasi infinite three-dimensional space. The plates do therefore not give rise to a boundary condition, which would complicate the mathematics considerably.
The degree of Taylor dispersion can be quantified by comparing the mean-squared displacement with and without shear flow. The mean-squared displacement $W(t) = < r^2 > (t) = < x^2 + y^2 + z^2 > (t)$ can be obtained from eq.(37) by multiplication of both sides with x^2, y^2, z^2 and xy, and integration. Subsequent partial integrations, as before, gives,

$$\begin{aligned}
\frac{d < x^2 > (t)}{d\,t} &= 2\,\dot\gamma\, < xy > (t) + 2\,D\;, \\
\frac{d < y^2 > (t)}{d\,t} &= 2\,D\;, \\
\frac{d < z^2 > (t)}{d\,t} &= 2\,D\;, \\
\frac{d < xy > (t)}{d\,t} &= \dot\gamma\, < y^2 > \;.
\end{aligned}$$

From the second equation we have $< y^2 > (t) = 2Dt$, so that from the last equation it is found that $< xy > (t) = \dot\gamma D t^2$. Substitution into the first equation and integration finally leads to,

$$W(t) \;=\; 6\,D\,t\,\left[1 + \tfrac{1}{9}\,(\dot\gamma\,t)^2\right]\;. \tag{38}$$

The mean-squared displacement thus increases with increasing shear rate. This shows that homogenization is more effective in the presence of flow, which is referred to Taylor dispersion. The effectiveness of flow-induced dispersion depends on the geometry. Taylor dispersion is originally investigated for flow through a cylindrical tube. The non-linear spatial dependence of

the flow velocity, the boundary conditions imposed by the tube's wall, and the initial condition (mostly a slab of a certain length with a constant concentration) do not allow for the evaluation of an analytical solution. Mathematical approximations are necessary, and various solutions valid in regions where dimensionless numbers are either large or small can be evaluated. This leads to diagrams where regions can be discerned where Taylor dispersion in such a cylindrical tube is effective (see, for example, Refs.[12, 13]).

5 Diffusion of Interacting Colloids

So far we considered the dynamics of single colloidal particles, which is relevant for the dynamics of colloids in dilute suspensions where the colloidal particles do not interact with each other. In this section we consider the effect on inter-colloidal interactions on particle dynamics and collective phenomena. We restrict ourselves here to translational diffusion, although rotational diffusion can be treated similarly. The translational diffusion equation will be generalized in subsection 5.1 to include interactions between colloidal particles, which equation of motion is referred to as the Smoluchowski equation. In subsection 5.2 we re-derive Fick's diffusion equation from the Smoluchowski equation, and obtain an expression for the concentration dependence of the translational diffusion coefficient. For very dilute suspensions this expression reduces to the Einstein result discussed in subsection 2.3. Finally, in subsection 5.3 it is shown that shear flow leads to a shear-induced mass flux resulting from interactions between the colloids, which in turn can give rise to an instability where the system "demixes" in regions with high and low shear rates.

There are two types of inter-colloidal interactions to be distinguished: direct interactions and hydrodynamic interactions. Direct interactions are due to, for example, excluded volumes (cores of colloidal particles can not interpenetrate), charges carried by the colloidal particles, and overlapping polymer brushes grafted to the surface of the particles (direct interactions between colloids are discussed in **lectures B4** and **F3**). *Hydrodynamic interactions* are mediated through the fluid in which the colloids are embedded: when a colloidal particle moves through the fluid, flow is induced that propagates to other colloidal particles and acts with a force on them (these interactions will be discussed in **lecture B3**). Although the principles of the theory presented below remain the same, the inclusion of hydrodynamic interactions complicates the analysis considerably. Furthermore, accurate analytical expressions for hydrodynamic interaction functions are only known on the two-particle level. For predictions at high concentrations one must therefore resort to computer simulations. For these two reasons, and the limited space and time for this lecture, we shall consider only direct interactions in the following.

5.1 The Smoluchowski equation

The Smoluchowski equation is an equation of motion for the pdf of the coordinates of an assembly of many colloids. This can be regarded as a generalization of Fick's diffusion equation, and is therefore some times also referred to as a *generalized diffusion equation*. In the following we consider spherical colloids, although the Smoluchowski equation can also be derived to include orientational degrees of freedom. Consider an assembly of N, possibly interacting, spherical colloidal particles. The probability function $P_N(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N, t)$ that is of interest here is related to the probability for the instantaneous realization of the positions of the centers-of-mass

$\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N\}$ of all N particles. To be more precise,

$P_N(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N, t)\, d\mathbf{r}_1\, d\mathbf{r}_2 \cdots d\mathbf{r}_N$ is the probability to find, at time t,
the center-of-mass of particle 1 inside the small volume $d\mathbf{r}_1$ centered at $\mathbf{r}_1$,
the center-of-mass of particle 2 inside the small volume $d\mathbf{r}_2$ centered at $\mathbf{r}_2$,
$\vdots$
and the center-of-mass of particle N inside the small volume $d\mathbf{r}_N$ centered at $\mathbf{r}_N$.

The rate-of-change of the N-particle pdf is simply the sum over all particles of terms that appear in the continuity equation (13) for a single sphere,

$$\frac{\partial}{\partial t} P_N \;=\; -\sum_{i=1}^{N} \nabla_i \cdot [\, \mathbf{v}_i\, P_N \,] \;,$$

where ∇_i is the gradient operator (6) with respect to $\mathbf{r}_i$, and $\mathbf{v}_i$ is the averaged velocity of particle i over time intervals that are large compared to the velocity relaxation time of a nanosecond, but small compared to times required for configurational changes. This velocity can again be obtained from force balance. The friction force (with the neglect of hydrodynamic interactions) is equal to $-\zeta\, \mathbf{v}_i$, the Brownian force generalizes to $-k_B T \nabla_i \ln P_N$, and there is a force $\nabla_i \Psi$ due to inter-colloidal interactions, where $\Psi \equiv \Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ is the total interaction energy of the assembly of the N colloidal spheres. Force balance implies that,

$$-\zeta\, \mathbf{v}_i - k_B T \nabla_i \ln P_N - \nabla_i \Psi \;=\; \mathbf{0} \;. \tag{39}$$

Substitution into the above continuity equation thus leads to what is known as *the Smoluchowski equation*,

$$\frac{\partial}{\partial t} P_N \;=\; D_0 \sum_{i=1}^{N} \nabla_i \cdot [\, \nabla_i P_N + \beta\, P_N\, \nabla_i \Psi \,] \;, \tag{40}$$

where $D_0 = k_B T/\zeta$ is Einstein's diffusion coefficient (the subscript "0" is added to indicate that this is the free diffusion coefficient, of a non-interacting sphere). This is the generalization of Fick's equation of motion (7) to an assembly of N interacting particles. As will be seen in the next subsection, Fick's equation of motion (7) still holds for concentrated systems, except that the diffusion coefficient is now concentration dependent.
The Smoluchowski equation describes the dynamics of collective phenomena resulting from interactions, both in equilibrium and out-of equilibrium. As such this equation is quite fundamental to colloid science.

5.2 The concentration dependence of Fick's diffusion coefficient

The single-particle pdf is by definition obtained by integration of the N-particle pdf, except for one of the position coordinates,

$$P(\mathbf{r}, t) \;=\; \int d\mathbf{r}_2 \int d\mathbf{r}_3 \cdots \int d\mathbf{r}_N\; P_N(\mathbf{r}, \mathbf{r}_2, \mathbf{r}_3, \cdots, \mathbf{r}_N, t) \;. \tag{41}$$

We will hereafter consider the concentration $\rho(\mathbf{r}, t)$ of colloids, which is defined as the number of colloids per unit volume at position $\mathbf{r}$ at time t. Since the single-particle pdf is proportional

to the local concentration, we can formulate an equation of motion for the concentration. As the pdf is normalized (it's integral is unity), while the integral of the concentration is equal to N by definition, we have,

$$\rho(\mathbf{r},t) \;=\; N\,P(\mathbf{r},t)\;. \tag{42}$$

Integration of both sides of the Smoluchowski equation (40) thus leads to,

$$\frac{\partial}{\partial\, t}\,\rho(\mathbf{r},t) \;=\; D_0\,\nabla\cdot\left[\,\nabla\rho(\mathbf{r},t) + \mathbf{\Delta}(\mathbf{r},t)\,\right]\;, \tag{43}$$

where Gauss's integral theorem has been used (see Appendix B), and where $\mathbf{\Delta}(\mathbf{r},t)$ describes the effect of interactions,

$$\mathbf{\Delta}(\mathbf{r},t) \;=\; \beta\,N\int d\mathbf{r}_2\cdots\int d\mathbf{r}_N\;P_N(\mathbf{r},\mathbf{r}_2,\mathbf{r}_3,\cdots,\mathbf{r}_N,t)\,\nabla\Psi(\mathbf{r},\mathbf{r}_2,\cdots,\mathbf{r}_N)\;.$$

Further progress can be made when assuming that the total interaction energy Ψ can be written as a pair-wise additive sum of interaction energies between two particles,

$$\Psi(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_N) \;=\; \sum_{i<j} V(|\,\mathbf{r}_i-\mathbf{r}_j\,|)\;, \tag{44}$$

where the two-particle interaction energy V is referred to as *the pair-interaction potential*, which is a function of the magnitude of the distance between two particles. The restriction $i<j$ in the summation avoids double counting of pairs of particles. Pair-wise additivity is exact for hard-sphere interactions, which are due to the impenetrability of the cores of two particles. It is also essentially exact when additional interactions are present, on top of the hard-core interactions, when these have a range that is small as compared to the core radius. In case of charged colloids, for example, with a small Debye length (see **lectures B4** and **F3**), pair-wise additivity is a very good approximation. For Debye lengths comparable or larger than the core radius, electric double layers of three particles can overlap simultaneously, leading to additional three-particle contributions to the pair-wise additive approximation in eq.(44).
Substitution of eq.(44) into the equation for $\mathbf{\Delta}$, and assuming identical particles, it is readily found that,

$$\mathbf{\Delta}(\mathbf{r},t) \;=\; \beta\,N\,(N-1)\int d\mathbf{r}\,'\;P_2(\mathbf{r},\mathbf{r}\,',t)\,\nabla V(|\,\mathbf{r}-\mathbf{r}\,'\,|)\;,$$

where,

$$P_2(\mathbf{r},\mathbf{r}\,',t) \;=\; \int d\mathbf{r}_3\int d\mathbf{r}_4\cdots\int d\mathbf{r}_N\;P_N(\mathbf{r},\mathbf{r}\,',\mathbf{r}_3,\mathbf{r}_4,\cdots,\mathbf{r}_N,t)\;,$$

is the two-particle pdf. If particles are not interacting, the two-particle pdf factorizes into two single-particle pdfs: $P_2(\mathbf{r},\mathbf{r}\,',t) = P(\mathbf{r},t)\,P(\mathbf{r}\,',t)$. For interacting particles a correcting factor must be introduced that accounts for correlations due to interactions,

$$P_2(\mathbf{r},\mathbf{r}\,',t) \;=\; P(\mathbf{r},t)\,P(\mathbf{r}\,',t)\,g(\mathbf{r},\mathbf{r}\,',t) \;=\; \frac{1}{N^2}\,\rho(\mathbf{r},t)\,\rho(\mathbf{r}\,',t)\,g(\mathbf{r},\mathbf{r}\,',t)\;, \tag{45}$$

where g is referred to as *the pair-correlation function*. Substitution into the above expression for $\mathbf{\Delta}(\mathbf{r},t)$ gives (note that $(N-1)/N = 1$ for $N \gg 1$),

$$\mathbf{\Delta}(\mathbf{r},t) \;=\; \beta\,\rho(\mathbf{r},t)\int d\mathbf{r}\,'\;\rho(\mathbf{r}\,',t)\,g(\mathbf{r},\mathbf{r}\,',t)\,\nabla V(|\,\mathbf{r}-\mathbf{r}\,'\,|)\;. \tag{46}$$

The range R_V of the pair-interaction potential is of the order of the size of the colloids, so that $V(|\,\mathbf{r}-\mathbf{r}\,'\,|)$ is essentially zero when the distance between $\mathbf{r}\,'$ and $\mathbf{r}$ is larger than R_V. The effective integration range in the above integral is therefore a region around $\mathbf{r}$ with a typical extent equal to R_V. The concentration $\rho(\mathbf{r}\,',t)$ varies only little (and approximately linear) over such small distances, so that we can Taylor expand (for notation conventions, see Appendix A),

$$\rho(\mathbf{r}\,',t) \;=\; \rho(\mathbf{r},t) + (\mathbf{r}\,'-\mathbf{r})\cdot\nabla\rho(\mathbf{r},t)\;. \tag{47}$$

Since the density evolves over time scales much larger than the time needed for the equilibration of the short-ranged part of the pair-correlation function ("short-ranged" since $|\,\mathbf{r}-\mathbf{r}\,'\,|< R_V$), the pair-correlation function can be taken equal to its equilibrium value at a concentration in between the points $\mathbf{r}$ and $\mathbf{r}\,'$,

$$g(\mathbf{r},\mathbf{r}\,',t) \;=\; g^{eq}(|\,\mathbf{r}-\mathbf{r}\,'\,|)\;\;,\;\;\text{at the concentration}\;\;,\;\;\rho=\rho\left(\tfrac{1}{2}\,(\mathbf{r}+\mathbf{r}\,'),t\right)\;,$$

where the superscript "eq" refers to "equilibrium". Note that the equilibrium pair-correlation function for an otherwise homogeneous system only depends on $|\,\mathbf{r}-\mathbf{r}\,'\,|$. There are theories available that give accurate expressions for the equilibrium pair-correlation function (in particular those based on the Ornstein-Zernike integral equation [1, 3]). Since for $|\,\mathbf{r}-\mathbf{r}\,'\,|< R_V$, the difference between $\rho\left(\frac{1}{2}\,(\mathbf{r}+\mathbf{r}\,'),t\right)$ and $\rho(\mathbf{r},t)$ is very small, the above expression can be Taylor expanded as,

$$\begin{aligned} g^{eq}(|\,\mathbf{r}-\mathbf{r}\,'\,|)_{|_{\rho=\rho\left(\frac{1}{2}\,(\mathbf{r}+\mathbf{r}\,'),t\right)}} \;&=\; g^{eq}(|\,\mathbf{r}-\mathbf{r}\,'\,|)_{|_{\rho=\rho(\mathbf{r},t)}} \\ &\quad+\frac{\partial}{\partial\,\rho}\,g^{eq}(|\,\mathbf{r}-\mathbf{r}\,'\,|)_{|_{\rho=\rho(\mathbf{r},t)}}\;\left[\,\rho\left(\tfrac{1}{2}\,(\mathbf{r}+\mathbf{r}\,'),t\right)-\rho(\mathbf{r},t)\,\right]\;. \end{aligned}$$

Next, the same gradient expansion of the density is made as above,

$$\rho\left(\tfrac{1}{2}\,(\mathbf{r}+\mathbf{r}\,'),t\right)-\rho(\mathbf{r},t) \;=\; \left\{\,\tfrac{1}{2}\,(\mathbf{r}+\mathbf{r}\,')-\mathbf{r}\,\right\}\cdot\nabla\rho(\mathbf{r},t) \;=\; \tfrac{1}{2}\;(\,\mathbf{r}\,'-\mathbf{r}\,)\cdot\nabla\rho(\mathbf{r},t)\;.$$

The pair-correlation function in the integral in eq.(46) can thus be written as (with $\mathbf{R}=\mathbf{r}\,'-\mathbf{r}$),

$$g(\mathbf{r},\mathbf{r}\,',t) \;=\; g^{eq}(|\,\mathbf{r}-\mathbf{r}\,'\,|)+\frac{1}{2}\,\frac{\partial\,g^{eq}(|\,\mathbf{r}-\mathbf{r}\,'\,|)}{\partial\,\rho(\mathbf{r},t)}\,\mathbf{R}\cdot\nabla\rho(\mathbf{r},t)\;,$$

where it is understood that the equilibrium pair-correlation function for the otherwise homogeneous system is evaluated at the local concentration $\rho(\mathbf{r},t)$. We thus find that,

$$\begin{aligned} \boldsymbol{\Delta}(\mathbf{r},t) \;&=\; -\,\beta\,\rho(\mathbf{r},t)\int d\mathbf{R}\;\left[\,\rho(\mathbf{r},t)+\mathbf{R}\cdot\nabla\rho(\mathbf{r},t)\,\right] \\ &\qquad\times\left[\,g^{eq}(R)+\tfrac{1}{2}\,\frac{\partial\,g^{eq}(R)}{\partial\,\rho(\mathbf{r},t)}\,\mathbf{R}\cdot\nabla\rho(\mathbf{r},t)\,\right]\,\frac{\mathbf{R}}{R}\,\frac{d\,V(R)}{d\,R}\;, \end{aligned}$$

where it is used that $\nabla V(R)=-(\mathbf{R}/R)\,dV(R)/dR$. Now using that integrals over odd products of $\mathbf{R}$ are zero due to symmetry, while,

$$\int d\mathbf{R}\;\mathbf{R}\mathbf{R} \;=\; \frac{4\,\pi}{3}\,R^4\,\hat{\mathbf{I}}\;,$$

with $\hat{\mathbf{I}}$ the identity tensor, it is readily found that,

$$\begin{aligned}\mathbf{\Delta}(\mathbf{r},t) &= -\beta\,[\,\nabla\rho(\mathbf{r},t)\,]\,\rho(\mathbf{r},t)\,\frac{2\,\pi}{3}\int_0^\infty dR\,R^3\,\left[\,2\,g^{eq}(R)+\rho(\mathbf{r},t)\,\frac{\partial\,g^{eq}(R)}{\partial\,\rho(\mathbf{r},t)}\,\right]\,\frac{d\,V(R)}{d\,R}\\ &= -\beta\,[\,\nabla\rho(\mathbf{r},t)\,]\,\frac{d}{d\,\rho(\mathbf{r},t)}\left\{\,\frac{2\,\pi}{3}\,\rho^2(\mathbf{r},t)\int_0^\infty dR\,R^3\,g^{eq}(R)\,\frac{d\,V(R)}{d\,R}\,\right\}\,.\end{aligned}$$

At this point it is convenient to introduce the well-known expression,

$$P^{eq}(\rho)\;=\;\rho\,k_B\,T-\frac{2\,\pi}{3}\,\rho^2\int_0^\infty dR\,R^3\,\frac{d\,V(R)}{d\,R}\,g^{eq}(R)\,, \tag{48}$$

for the equilibrium pressure P^{eq} of a homogeneous system with concentration ρ [14]. The first term on the right hand-side is the ideal gas pressure, the second term accounts for contributions to the pressure resulting from inter-colloidal interactions. This allows us to rewrite,

$$\mathbf{\Delta}(\mathbf{r},t)\;=\;\beta\,[\,\nabla\rho(\mathbf{r},t)\,]\,\frac{d}{d\,\rho(\mathbf{r},t)}\,[\,P^{eq}(\rho(\mathbf{r},t))-\rho(\mathbf{r},t)\,k_B\,T\,]\;.$$

Note that this can also be written as,

$$\mathbf{\Delta}(\mathbf{r},t)\;=\;\beta\,\nabla\,[\,P^{eq}(\rho(\mathbf{r},t))-\rho(\mathbf{r},t)\,k_B\,T\,]\;. \tag{49}$$

Substitution into eq.(43) thus finally leads to,

$$\frac{\partial}{\partial\,t}\,\rho(\mathbf{r},t)\;=\;\nabla\cdot[\,D(\rho(\mathbf{r},t))\,\nabla\rho(\mathbf{r},t)\,]\;, \tag{50}$$

where *the collective diffusion coefficient* is equal to,

$$D(\rho)\;=\;\beta\,D_0\,\frac{d\,P^{eq}(\rho)}{d\rho}\;. \tag{51}$$

The only restriction for the validity of this diffusion equation is that the density is essentially a linear function of position over distance of the order of the range R_V of the pair-interaction potential. For larger gradients in the concentration, higher order derivatives in the spatial Taylor expansions must be accounted for. This is typically the case for sharp interfaces between two phases with different colloid concentrations.
In case the variations in density are not too large, so that $\rho(\mathbf{r},t)=\bar{\rho}+\delta\rho(\mathbf{r},t)$, with $|\,\delta\rho(\mathbf{r},t)/\bar{\rho}\ll 1$, eq.(50) can be linearized with respect to $\delta\rho(\mathbf{r},t)$, which leads to,

$$\frac{\partial}{\partial\,t}\,\rho(\mathbf{r},t)\;=\;D(\bar{\rho})\,\nabla^2\rho(\mathbf{r},t)\;.$$

This reproduces Fick's law (7), but now with a concentration dependent diffusion coefficient. For the more general case of large concentration variations (but still with sufficiently small spatial gradients), eq.(50) should be used, which is generally a non-linear equation of motion due to the concentration dependence of the diffusion coefficient.
It should be mentioned that the expression (51) for the diffusion coefficient is only accurate when hydrodynamic interactions are not important. This is the case, for example, for charged colloids with a Debye length that is larger than the core radius (for which, however, the pair-wise additivity approximation in eq.(44) is questionable), and where the fraction of the volume occupied by the cores of the colloids is small.

5.3 A shear-induced instability

Here we repeat the above analysis, but now for a system that is subjected to flow. As will be seen, this leads to a shear-rate dependent diffusion coefficient, and more importantly to a mass flux that is induced by spatial gradients in the shear rate. This shear-induced mass flux from regions of high shear rate to regions of low shear rate is entirely due to direct interactions between the colloids, and was first discussed in Refs.[15, 16] on a semi-empirical basis, as the origin of the shear-induced mass flux was not clear. The present analysis is taken from Refs.[17, 18, 19] (where the first two references are overview articles), which elucidates the microscopic origin of the shear-induced mass transport. The instability caused by this mass transport can be understood intuitively as follows. Consider a Couette geometry, where the sample is contained in between two concentric cylinders, of which the inner- or outer-cylinder is rotating to induce flow. The shear rate near the outer cylinder is smaller as compared to the inner cylinder. There is thus mass transport towards the outer cylinder. The increase of concentration near the outer cylinder leads to an increase of the local stress, provided that the viscosity is an increasing function of concentration. The response of the system is to decrease the stress by lowering the local shear rate. This amplifies spatial gradients of the shear rate, and leads to an enhancement of mass transport towards the outer cylinder. This is the self-amplifying mechanism that renders the system unstable. A stationary state is reached when diffusive mass transport due to concentration gradients is compensated by the mass transport resulting from spatial gradients in the shear rate.
It will be shown that this mass flux can lead to an instability in glassy systems, where the stationary state consists of two regions: a flowing region with a constant non-zero shear rate, and a region that does not flow. This instability is referred to as *the Shear-gradient Concentration Coupling* (SCC-) instability, which should not be confused with the classic gradient-banding instability.

5.3.1 The advection-diffusion equation

The difference between the force-balance equation (39) is that there is now a flow velocity $\mathbf{u}(\mathbf{r},t)$ which can be represented by the so-called *velocity-gradient tensor* $\hat{\mathbf{\Gamma}}$,

$$\mathbf{u}(\mathbf{r},t) \;=\; \dot{\gamma}\,\hat{\mathbf{\Gamma}}\cdot\mathbf{r}\;, \tag{52}$$

where $\dot{\gamma}$ is the shear rate, which measures the magnitude of spatial gradients in the flow velocity. For the special case where,

$$\hat{\mathbf{\Gamma}} \;=\; \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\;,$$

this the flow that has been considered in subsection 4.3 on Taylor dispersion. Here, this representation is understood to be the local flow. The direction and magnitude of the flow will later be allowed to vary with position. Since the colloidal spheres are dragged along by the local flow, the force-balance equation (39) now reads,

$$-\zeta\;(\,\mathbf{v}_i-\mathbf{u}(\mathbf{r_i},t)\,)-k_BT\nabla_i\ln P_N-\nabla_i\Psi\;=\;\mathbf{0}\;,$$

where, as before, hydrodynamic interactions are neglected. The Smoluchowski equation therefore reads,

$$\frac{\partial}{\partial t}P_N \;=\; D_0\sum_{i=1}^{N}\nabla_i\cdot\left[\,\nabla_i P_N+\beta\,P_N\,\nabla_i\Psi\,\right]-\sum_{i=1}^{N}\nabla_i\cdot\left[\,P_N\,\mathbf{u}(\mathbf{r}_i,t)\,\right]\;, \tag{53}$$

Using the connection (41,42) between P_N and the concentration $\rho(\mathbf{r},t)$, it is found by integration, just like in the previous subsection, that,

$$\frac{\partial}{\partial t}\,\rho(\mathbf{r},t) \;=\; D_0\,\nabla\cdot\left[\,\nabla\rho(\mathbf{r},t)+\mathbf{\Delta}(\mathbf{r},t)\,\right]-\nabla\cdot\left[\,\rho(\mathbf{r},t)\,\mathbf{u}(\mathbf{r},t)\,\right]\;, \tag{54}$$

where $\mathbf{\Delta}(\mathbf{r},t)$ is given in eq.(46), and where the pair-correlation function is defined in eq.(45). The pair-correlation function is now the sum of the pair-correlation without shear flow and a contribution due to the shear flow. For the small distances $|\,\mathbf{r}-\mathbf{r}\,'|< R_V$ of interest in the above integral, the shear-distortion of the pair-correlation function is affine. Furthermore, similar to the previous subsection, the temporal evolution of the concentration and shear rate variations are much slower that the relaxation time of the pair-correlation function for distances less than R_V, so that the pair-correlation function adjusts itself essentially instantaneously to its local stationary form. The time dependence of the pair-correlation function is therefore entirely due to the time dependence of the local density and shear rate. Hence (for notation, see Appendix A),

$$g(\mathbf{r},\mathbf{r}\,',t) \;=\; g^{eq}(\mathbf{r},\mathbf{r}\,',t)+g_0(\mathbf{r},\mathbf{r}\,',t)\;+\;\frac{\mathbf{r}-\mathbf{r}\,'}{|\,\mathbf{r}-\mathbf{r}\,'|}\cdot\hat{\mathbf{E}}\cdot\frac{\mathbf{r}-\mathbf{r}\,'}{|\,\mathbf{r}-\mathbf{r}\,'|}\,g_1(\mathbf{r},\mathbf{r}\,',t)\;, \tag{55}$$

where, as before, g^{eq} is the equilibrium pair-correlation function in the absence of shear flow, g_0 is the isotropic shear-induced distortion, and g_1 characterizes the anisotropic affine distortion of the pair-correlation function. Furthermore, $\hat{\mathbf{E}}$ is the symmetric part of the velocity-gradient tensor $\hat{\mathbf{\Gamma}}$,

$$\hat{\mathbf{E}} \;=\; \frac{1}{2}\begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix}\;.$$

The further analysis is very much along the lines of the previous subsection, but now with the shear rate as an additional variable to the concentration, and a different form of the pair-correlation function.

The shear rate and colloid density vary only little over distances less than R_V, so that the density $\rho(\mathbf{r}\,',t)$ appearing in the integral in the above expression for $\mathbf{\Delta}(\mathbf{r},t)$ can be Taylor expanded around $\mathbf{r}$ to leading order in gradients, as in eq.(47). For the same reason, the various contributions to the pair-correlation function in eq.(55) are approximately equal to those in a homogeneous system, with the concentration and shear rate taken in between the positions $\mathbf{r}$ and $\mathbf{r}\,'$. For example,

$$\begin{aligned} g_0(\mathbf{r},\mathbf{r}\,',t) \;&=\; \bar{g}_0(|\,\mathbf{r}-\mathbf{r}\,'|,t)\;,\\ &\qquad\text{at the concentration}\quad \bar{\rho}=\rho\left(\tfrac{1}{2}\,(\mathbf{r}+\mathbf{r}\,'),t\right)\;,\\ &\qquad\text{and at the shear rate}\quad \bar{\dot{\gamma}}=\dot{\gamma}\left(\tfrac{1}{2}\,(\mathbf{r}+\mathbf{r}\,'),t\right)\;, \end{aligned}$$

and similar for g^{eq} and g_1. The overbar on $\bar{g}_0$ is used to indicate that this is the correlation function of a homogeneous system with density $\bar{\rho}$ and shear rate $\bar{\dot{\gamma}}$. Repeating he mathematical steps of the previous subsection, but now including the shear rate, leads to (again with $\mathbf{R} = \mathbf{r} - \mathbf{r}\,'$),

$$g_0(\mathbf{r}, \mathbf{r}\,', t) \;=\; \bar{g}_0(R,t) + \frac{1}{2}\,\frac{\partial \bar{g}_0(R\,|\,\rho,\dot{\gamma})}{\partial \rho}\;\mathbf{R}\cdot\nabla\rho + \frac{1}{2}\,\frac{\partial \bar{g}_0(R\,|\,\rho,\dot{\gamma})}{\partial \dot{\gamma}}\;\mathbf{R}\cdot\nabla\dot{\gamma}\,,$$

and similar for g^{eq} (for which there is no shear-rate dependence) and g_1. Here, ρ and $\dot{\gamma}$ are now understood to denote the local density and shear rate (hereafter we omit the overbar notation, and the explicit $\mathbf{r}$- and t-dependence of ρ and $\dot{\gamma}$). Substitution of these results into eq.(46), and performing the angular integrations as in the previous subsection, leads to,

$$\begin{aligned}\mathbf{\Delta}(\mathbf{r},t) \;=\; \beta\,\Bigg\{ &\left[\frac{\partial P_0(\rho,\dot{\gamma})}{\partial\rho} - k_B T\right]\nabla\rho + \frac{\partial P_0(\rho,\dot{\gamma})}{\partial\dot{\gamma}}\,\nabla\dot{\gamma} \\ &+\frac{\partial P_1(\rho,\dot{\gamma})}{\partial\rho}\,\hat{\mathbf{E}}\cdot\nabla\rho + \frac{\partial P_1(\rho,\dot{\gamma})}{\partial\dot{\gamma}}\,\hat{\mathbf{E}}\cdot\nabla\dot{\gamma}\Bigg\}\,,\end{aligned}$$

up to leading order in spatial gradients, where,

$$P_0(\rho,\dot{\gamma}) \;=\; \rho\,k_B T - \frac{2\,\pi}{3}\,\rho^2\int_0^\infty dR\;R^3\,\frac{dV(R)}{dR}\;\left[\,g^{eq}(R\,|\,\rho) + \bar{g}_0(R\,|\,\rho,\dot{\gamma})\,\right]\,, \tag{56}$$

and,

$$P_1(\rho,\dot{\gamma}) \;=\; -\frac{4\,\pi}{15}\,\rho^2\int_0^\infty dR\;R^3\,\frac{dV(R)}{dR}\;\bar{g}_1(R\,|\,\rho,\dot{\gamma})\,.$$

Note that the above expression for $\mathbf{\Delta}(\mathbf{r},t)$ can also be written as,

$$\mathbf{\Delta}(\mathbf{r},t) \;=\; \beta\,\Big\{\,\nabla\left[\,P_0(\rho,\dot{\gamma}) - \rho\,k_B T\,\right] + \hat{\mathbf{E}}\cdot\nabla P_1(\rho,\dot{\gamma})\,\Big\}\,, \tag{57}$$

where the gradient operator acts on the position dependence of ρ and $\dot{\gamma}$, which dependence is not denoted explicitly for convenience. The only component of $\Delta(\mathbf{r},t)$ of interest here is the component acting along the gradient direction, which is the contribution $\beta\,\nabla\,[P_0 - \rho\,k_B\,T]$. We thus finally arrive at the equation of motion for the density from eq.(54),

$$\frac{\partial}{\partial\,t}\,\rho(\mathbf{r},t) \;=\; \nabla\cdot\left[\,D(\rho,\dot{\gamma})\nabla\rho(\mathbf{r},t)\,\right] + \nabla\cdot\left[\,\xi(\rho,\dot{\gamma})\,\nabla\dot{\gamma}\,\right] - \nabla\cdot\left[\,\rho(\mathbf{r},t)\,\mathbf{u}(\mathbf{r},t)\,\right]\,, \tag{58}$$

where the concentration and shear-rate dependence of the transport coefficients D and ξ are understood to be the local, $\mathbf{r}$- and t-dependent quantities. The two transport coefficients are equal to,

$$\begin{aligned} D(\rho,\dot{\gamma}) \;&=\; \beta\,D_0\,\frac{\partial\,P_0(\rho,\dot{\gamma})}{\partial\,\rho}\,, \\ \xi(\rho,\dot{\gamma}) \;&=\; \beta\,D_0\,\frac{\partial\,P_0(\rho,\dot{\gamma})}{\partial\,\dot{\gamma}}\,. \end{aligned} \tag{59}$$

We will refer to the transport coefficient ξ as *the shear-gradient coefficient*. The quantity P_0 is nothing but the equilibrium pressure in eq.(48), except that the pair-correlation function is now

replaced by the isotropic contribution to the shear-distorted pair-correlation function. There is mass transport from regions of high shear rate towards regions of low shear rate, provided that $\xi > 0$, which is the case for repulsive interactions.
The new feature in eq.(58) is that there is a contribution to mass transport due to spatial gradients of the shear rate. Such shear-induced mass transport can give rise to an instability, as discussed in the introduction to this subsection. In the following we consider suspensions of hard spheres in the glass state, for which the pair-interaction potential is either infinite when cores overlap, or is zero otherwise. For such hard-core interactions the integral in eq.(56) for P_0 can be evaluated in terms of the contact value of the sum of the two pair-correlation functions appearing in the integrand, that is, the value where the distance between two colloids is equal to $2a$, with a the radius of the cores,

$$g^c_{iso}(\rho, \dot{\gamma}) \;=\; g^{eq}(R = 2a \,|\, \rho) + \bar{g}_0(R = 2a \,|\, \rho, \dot{\gamma}) \;,$$

where the superscript "c" stands for "contact value", while the subscript "iso" stands for "isotropic part" of the total pair-correlation function. The evaluation of the integral can be done by introducing the so-called cavity function $y = g\, \exp\{+\beta\, V\}$ [14]. This function has the same contact value as the pair-correlation function. Since $\exp\{-\beta V\} dV/dR = -k_B T d \exp\{-\beta V\}/dR = -k_B T \delta(R - 2a)$, with δ the delta-distribution, it follows that,

$$P_0(\rho, \dot{\gamma}) \;=\; \rho\, k_B T + \frac{2\,\pi}{3}\, (2a)^3\, \rho^2\, k_B T\, g^c_{iso}(\rho, \dot{\gamma}) \;.$$

In terms of the dimensionless concentration,

$$\phi \;\; = \frac{4\pi}{3}\, a^3\, \rho \;,$$

the volume fraction of colloidal spheres, and the dimensionless shear rate, the so-called *Peclet number* (also denoted by Pe),

$$\widetilde{\dot{\gamma}} \;=\; \dot{\gamma}\, a^2/D_0 \;.$$

it follows from eq.(59) that,

$$D_{eff}(\phi, \widetilde{\dot{\gamma}}) \;=\; D_0 \left[1 + 4\, \frac{\partial}{\partial \phi} \left\{ \phi^2\, g^c_{iso}(\phi, \widetilde{\dot{\gamma}}) \right\} \right] \;,$$

and the shear-gradient coefficient is equal to,

$$\xi(\phi, \widetilde{\dot{\gamma}}) \;=\; 18\, \eta_0\, \frac{D_0}{k_B T}\, \phi^2\, \frac{\partial g^c_{iso}(\phi, \widetilde{\dot{\gamma}})}{\partial \widetilde{\dot{\gamma}}} \;.$$

The explicit density and shear-rate dependence of the contact value of the pair-correlation function must be obtained from Brownian Dynamics simulations (see **lecture D2**). It turns out that a quite accurate representation of the simulation results is [19],

$$g^c_{iso}(\phi, \widetilde{\dot{\gamma}}) \;=\; g^{eq,\,c}(\phi) + A\, \widetilde{\dot{\gamma}}^{\,m} \left(1 - \frac{\phi}{\phi_m} \right)^{-s} \;,$$

where $A = 0.0140$, $s = 2.525$, and $m = 0.43 + 5.26\,\phi - 8.80\,\phi^2$, while $\phi_m = 0.64$ is the maximum random packing fraction, at which flow is not possible anymore. Furthermore $g^{eq,c}$

is the contact value of the equilibrium pair-correlation function, for which accurate expressions exist, like the Carnahan-Starling approximation [20],

$$g^{eq,c}(\phi) \;=\; \frac{2-\phi}{2(1-\phi)^3} \;.$$

The above representation for the contact value of the pair-correlation function is accurate for small shear rates, such that $\widetilde{\dot{\gamma}} < 1$.
Now the concentration and shear-rate dependence of the transport coefficients in the advection-diffusion equation (58) is specified, an equation of motion for the flow velocity $\mathbf{u}(\mathbf{r},t)$ must be constructed in order to fully analyze the SCC-instability.

5.3.2 The equation of motion for the suspension flow velocity

An equation of motion is constructed from what is known through experiments on glassy hard-sphere colloidal systems. A general equation of motion for the flow velocity is *the Navier-Stokes equation* (see **lecture C6**), which states that,

$$\rho_m \frac{\partial\, \mathbf{u}(\mathbf{r},t)}{\partial\, t} \;=\; \nabla\cdot\boldsymbol{\Sigma}(\mathbf{r},t) \;,$$

where ρ_m is the specific mass of the suspension and $\boldsymbol{\Sigma}$ is stress tensor, which is equal to ("T" stands for "transposition"),

$$\boldsymbol{\Sigma} \;=\; \boldsymbol{\Sigma}_{yield} + \left(\,\eta - \kappa\,\nabla^2\right)\,\left\{\,\nabla\mathbf{u} + (\nabla\mathbf{u})^T\,\right\} \;,$$

where $\boldsymbol{\Sigma}_{yield}$ is the yield-stress tensor, which is non-zero for hard-sphere volume fractions larger than the glass-transition volume fraction $\phi_g = 0.58$, and where η is the shear-viscosity. Spatial variations in the pressure and normal stresses are of no significance for the flow profiles under consideration here, which are therefore omitted. There is a non-standard contribution in the above expression for the stress tensor, of which the amplitude is equal to the so-called *shear-curvature viscosity* $\kappa > 0$. This non-local contribution to the stress is essential in an analysis of banded flow profiles, as it describes stresses that are due to the large spatial gradients in the flow velocity that exist within the interface between the bands [21].
The temporal evolution of the concentration is much slower than the adjustment of flow to the local concentration. The flow velocity thus quasi-instantaneously adjusts to the concentration, that is, flow is *enslaved* by concentration. On a time scale that is large compared to the adjustment time of the flow velocity, and small compared to the typical times for significant changes of concentration due to the SCC-instability, the time derivative in the Navier-Stokes equation can be set equal to zero. Hence,

$$0 \;=\; \nabla\cdot\left[\,\boldsymbol{\Sigma}_{yield} + \left(\,\eta - \kappa\,\nabla^2\right)\,\left\{\,\nabla\mathbf{u} + (\nabla\mathbf{u})^T\,\right\}\,\right] \;, \tag{60}$$

As for the advection-diffusion equation, the transport coefficients η and κ are functions of the local concentration and shear rate, for which we will use experimental results.
The shear viscosity of glassy systems of hard spheres decreases with increasing shear rate once the applied stress is larger than the yield stress. There is no Newtonian plateau for glasses, contrary to fluids, where the viscosity is independent of the shear rate for sufficiently small shear rates. Shear thinning immediately sets in when flow is induced, which is reasonably well

described by a so-called Herschel-Bulkley form of the stress, which predicts that the viscosity varies like $\sim \widetilde{\dot{\gamma}}^{-1/2}$ [22, 23, 24],

$$\eta(\phi,\widetilde{\dot{\gamma}}) \;=\; 15\,\frac{a^2}{D_0}\,\Sigma_{yield}(\phi)\,\left(1-\frac{\phi}{\phi_m}\right)^{1/2}\widetilde{\dot{\gamma}}^{-1/2}\;.$$

The concentration dependence of the yield stress of hard-sphere glasses is $\sim (1-\Phi)^{-p}$, where p is reported to vary between 1 and 3, while the prefactor is equal to $k_BT/(100a^3)$ [22, 25]. We will use the same expression for the yield stress as in Ref.[22],

$$\Sigma_{yield}(\phi) \;=\; \frac{k_BT}{100\,a^3}\,\left(1-\frac{\phi}{\phi_m}\right)^{-3}\;,$$

which is understood to act along the flow direction. The shear-curvature viscosity exhibits a similar shear-rate and concentration dependence as the shear viscosity,

$$\kappa(\phi,\widetilde{\dot{\gamma}}) \;=\; 15\,\frac{a^2}{D_0}\,\frac{\kappa_0}{\eta_0}\,\Sigma_{yield}(\Phi)\,\left(1-\frac{\phi}{\phi_m}\right)^{1/2}\widetilde{\dot{\gamma}}^{-1/2}\;,$$

where η_0 is the shear viscosity of the solvent, and κ_0 measures the magnitude of the shear-curvature viscosity. In case the stationary state is a shear-banded state, where two regions (the "bands") coexist, the interface thickness is of the order,

$$d_{int} \;=\; \sqrt{\frac{\kappa_0}{\eta_0}}\;. \tag{61}$$

This completes the construction of the equation of motion for the flow velocity, and thereby the two coupled equations of motion (58,60) that are necessary for the further analysis of the SCC-instability.

5.3.3 Stability and flow profiles

Let us first ask for the stability of a uniform system, that is, we ask for the combination of concentrations and shear rates where a uniform system with concentration ρ_0 and shear rate $\dot{\gamma}_0$ will evolve towards a stationary non-uniform state upon applying initially arbitrary small perturbations $\delta\rho$ of the concentration and δu of the flow velocity. We consider a flow in the x-direction with gradients in density and flow velocity only in the y-direction. The applied shear stress is supposed to be larger than the yield stress. Substitution of $\rho = \rho_0+\delta\rho$ and $u = \dot{\gamma}_0 y + \delta u$ into the advection-diffusion equation (58) and the Navier-Stokes equation (60), and linearization with respect to the small perturbations $\delta\rho$ and δu gives,

$$\begin{aligned}\frac{\partial\delta\rho}{\partial t} \;&=\; D_{eff}\frac{\partial^2\delta\rho}{\partial y^2}+\xi\frac{\partial^3\delta u}{\partial y^3}\;,\\ 0 \;&=\; \frac{\partial\sigma}{\partial\dot{\gamma}_0}\,\frac{\partial^2\delta u}{\partial y^2}+\frac{\partial\sigma}{\partial\rho_0}\frac{\partial\delta\rho}{\partial y}-\kappa\frac{\partial^4\delta u}{\partial y^4}\;,\end{aligned} \tag{62}$$

where D_{eff}, ξ, η and κ are evaluated at ρ_0 and $\dot{\gamma}_0$, and where,

$$\sigma \;=\; \Sigma_{yield}(\rho_0)+\dot{\gamma}_0\,\eta(\rho_0,\dot{\gamma}_0)\;,$$

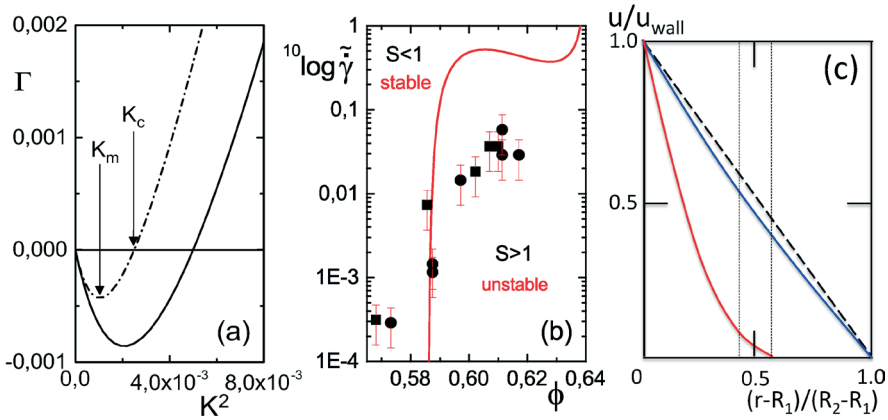


Fig. 7: *(a) The dimensionless (negative) growth rate versus the squared dimensionless wave vector, for $S = 2$ and for $C = 200$ and 400. (b) The stability diagram in the shear-rate versus concentration plane. The red curve marks the shear rates and concentrations where $S = 1$. The points are experimental data for PMMA spheres with a radius of $138\,nm$ (the circles) and $150\,nm$ (the squares). These data are taken from Ref.[22], with courtesy of Wilson Poon. (c) Flow profiles for a narrow-gap Couette geometry. The velocity is normalized to the velocity u_{wall} of the rotating inner cylinder. R_1 and R_2 are the radii of the inner and outer cylinder, respectively. The dashed line is for a stable system without shear-gradient induced mass transport, the solid blue line for a stable system including mass transport, and the red solid curve for an unstable system. The two vertical lines indicate the thickness of the interface.*

is the shear-stress of the initially homogeneously sheared suspension. The time dependence of the perturbations is exponential due to the linearization, while the exponents for the density and flow velocity are the same, as the velocity is enslaved by the concentration. Hence, for sinusoidal spatial perturbations (with $i = \sqrt{-1}$),

$$\begin{aligned} \delta\rho &= \delta\rho_0 \, \exp\{iky - \lambda(k)t\} \, , \\ \delta u &= \delta u_0 \, \exp\{iky - \lambda(k)t\} \, . \end{aligned} \tag{63}$$

The variable k is referred to as *the wave vector*, which has the following physical meaning. In real-space notation, the spatial variation is $\sim \sin\{k\,y\} = \sin\{2\pi\,y/\Lambda\}$, where $\Lambda = 2\pi/k$ is the wavelength of the sinusoidal variation. Thus, smooth spatial variations (large wavelengths Λ) correspond to small values of the wave vector k. If $\lambda(k) < 0$, eq.(63) implies that sinusoidal variations with a wavelength $\Lambda = 2\pi/k$ will grow in amplitude, that is, such spatial variations are unstable. Substitution of the representation (63) into eq.(62) leads to *the dispersion relation*,

$$\lambda(k) \;=\; k^2 \left[D_{eff} - \xi \, \frac{d\sigma}{d\rho} \left\{ \frac{d\sigma}{d\dot{\gamma}} + \kappa \, k^2 \right\}^{-1} \right] \, . \tag{64}$$

This dispersion relation is most conveniently written in dimensionless form as,

$$\Gamma \;=\; K^2 \left[1 - \frac{S}{1 + C\,K^2} \right] \, ,$$

where $\Gamma = \lambda(k)\, a^2/D_{eff}$ (with a the radius of the colloidal spheres), $K = k\, a$ is the dimensionless wave vector, S is equal to,

$$S \equiv \frac{\xi}{D_{eff}} \frac{d\sigma/d\rho}{d\sigma/d\dot{\gamma}} , \tag{65}$$

and $C = \kappa/(a^2\, d\sigma/d\dot{\gamma})$. The system is unstable when $S > 1$, since then $\Gamma < 0$, which is the reason that S is referred to as *the stability factor*. The dimensionless negative growth rate Γ/K^2 is plotted in Fig.7a as a function of K^2 for the typical values of $S = 2$, and $C = 200$ (solid curve) and $C = 400$ (dashed-dotted curve). Those wave vectors where $\Gamma < 0$ are unstable, while larger wave vectors (corresponding to large spatial gradients, as discussed above) remain stable. The stabilization of sharp spatial variations, corresponding to large wave-vectors, is solely due to the non-local stress contribution, that is, the non-zero positive value of the shear-curvature viscosity κ. Without this non-local stress contribution all wave vectors would be unstable, with an arbitrary large growth rate for large wave vectors, which is unphysical. The larger the shear-curvature viscosity (corresponding to larger values of C), the smaller the critical wave vector beyond which spatial variations are stabilized.

The critical wave vector k_c for which $\lambda)k_c) = 0$, beyond which spatial variations are stable, follows from the dispersion relation (64) as,

$$k_c = \sqrt{\frac{d\sigma/d\dot{\gamma}}{\kappa}\,(S-1)} \quad , \quad (S > 1) , \tag{66}$$

while the fastest growing mode is the one with the wave vector k_m for which $d\lambda/dk = 0$, and hence,

$$k_m = \sqrt{\frac{d\sigma/d\dot{\gamma}}{\kappa}\left(\sqrt{S}-1\right)} \quad , \quad (S > 1) . \tag{67}$$

This is the wave vector where Γ attains its minimum in Fig.7a.

The stability criterion $S < 1$ together with the explicit forms of the transport coefficients discussed before allows to construct a stability diagram. This diagram marks the combinations of shear rates and concentrations where the system turns from (meta-) stable to unstable, where the stability factor in eq.(65) is unity. The curve where $S = 1$ in the shear rate versus concentration plane is shown in Fig.7b. The experimental data points for suspensions of PMMA spheres of two different radii are taken from Ref.[22]. Note that the hard-sphere glass is predicted to be unstable at low shear rates, which is in accordance with the experiments. In view of our neglect of hydrodynamic interactions, the correspondence between theory and experiment is quite satisfactory.

The two coupled equations of motion (58,60) can be solved numerically. This is done for a Couette geometry with a small gapwidth: the ratio of the radius of the inner cylinder and the outer cylinder is chosen as 0.98, with a gapwidth equal to 1000 times the radius of the colloidal spheres. A numerical solution is only possible when the non-local stress distribution $\sim \kappa$ is included. When this contribution is not included, any numerical algorithm is unstable because of the arbitrary fast growth of sharp spatial gradients. The interface thickness d_{int} in eq.(61) is probably much larger than the size of the colloidal particles, so that the choice $d_{int} = 100\, a$ seems reasonable (with a the radius of the colloidal spheres).

The two blue-colored flow curves in Fig.7c correspond to a stable flow ($S < 1$) without the shear-induced mass-transport contribution (the dashed line), and with the shear-induced mass

transport. Shear-induced mass transport leads to a slightly more curved flow profile. The lower solid red line is for an unstable flow, where mass transport is sufficiently strong as compared to diffusion to render the system SCC-unstable. The two vertical lines indicate the width of the interface in eq.(61). As can be seen, there is a non-flowing band, where the local stress is lower than the yield stress pertaining to the local concentration. Outside the interface the shear rate within the flowing band is essentially constant. Such stationary banded states have indeed been observed in suspensions of hard spheres in the glass state, where however, a cone-plate geometry has been used [22].

Appendices

A The Contraction Notation

A dot between two vectors $\mathbf{a}$ and $\mathbf{b}$ implies the operation where the indices of the vectors are set equal and the sum over indices is taken: $\mathbf{a} \cdot \mathbf{b} = \sum_{n=1}^{3} a_n\, b_n$, which is *the inner product* of two vectors. *The dyadic product* $\mathbf{ab}$ of two vectors defines the tensor $\mathbf{M}$ with elements $M_{mn} = a_n b_m$. Two dots in between two tensors $\mathbf{M}$ and $\mathbf{K}$ is a generalization of the inner product of two vectors, where the adjacent and next adjacent indices are set equal and summed over: $\mathbf{M} : \mathbf{K} = \sum_{m,n=1}^{3} M_{mn} K_{nm}$. The notation $(\mathbf{r}' - \mathbf{r})(\mathbf{r}' - \mathbf{r}) : \nabla\nabla P(\mathbf{r}, t)$ in eq.(5) thus stands for $\sum_{m,n=1}^{3} (\mathbf{r}' - \mathbf{r})_m (\mathbf{r}' - \mathbf{r})_n \nabla_n \nabla_m P(\mathbf{r}, t)$, where ∇_n is the nth component of the vector operator in eq.(6). The same notation is used in eq.(18): $\nabla \cdot \mathbf{D} \cdot \nabla(\cdots)$ stands for $\sum_{m,n=1}^{3} \nabla_m\, D_{mn}\, \nabla_n (\cdots)$.
The summations over indices are commonly referred to as *contractions*, for which the "dot-notation" is a convenient short-hand notation.

B Gauss's and Stokes's Integral Theorems and Partial Integration

Gauss's integral theorem states that (with $d\mathbf{S} = \hat{\mathbf{n}}\, dS$, and $\mathbf{F} = (F_x, F_y, F_z)$ a vector function),

$$\int_V d\mathbf{r}\ \nabla \cdot \mathbf{F}(\mathbf{r}) \ = \ \oint_{\partial V} d\mathbf{S} \cdot \mathbf{F}(\mathbf{r})\ ,$$

where ∇ is again the gradient operator in eq.(6), V is the volume over which the integration extends, and ∂V is the closed surface of V. Note that by definition,

$$\nabla \cdot \mathbf{F} \ = \ \frac{\partial}{\partial x} F_x + \frac{\partial}{\partial y} F_y + \frac{\partial}{\partial z} F_z\ ,$$

which is referred to as *the divergence* (of $\mathbf{F}$). This is the inner product of the nabla operator in eq.(6) with the vector function $\mathbf{F}$.
Let V be a sphere of radius R. In the limit that $R \to \infty$, the integration range extends over the entire three-dimensional space, The spherical surface ∂V tends to infinity, with a surface area $\sim R^2$. If the integrand $|\,\mathbf{F}\,|$ tends to zero faster than $\sim 1/R^2$, the surface integral tends to zero, so that,

$$\int d\mathbf{r}\ \nabla \cdot \mathbf{F}(\mathbf{r}) \ = \ 0\ ,$$

where the integral is understood to range over the entire three-dimensional space.
For a scalar function $f(\mathbf{r})$ and a vector function $\mathbf{G}(\mathbf{r})$ we have the following identity,

$$\int d\mathbf{r}\; f(\mathbf{r})\nabla\cdot\mathbf{G}(\mathbf{r}) \;=\; \int d\mathbf{r}\;\nabla\cdot[\,f(\mathbf{r})\,\mathbf{G}(\mathbf{r})\,] - \int d\mathbf{r}\;\mathbf{G}(\mathbf{R})\cdot\nabla f(\mathbf{r})\;.$$

Using the above result with $\mathbf{F} = f\,\mathbf{G}$, we find that,

$$\int d\mathbf{r}\; f(\mathbf{r})\nabla\cdot\mathbf{G}(\mathbf{r}) \;=\; -\int d\mathbf{r}\;\mathbf{G}(\mathbf{r})\cdot\nabla f(\mathbf{r})\;,$$

This represents a single partial integration in three dimensions, which is used in subsections 4.1 and 4.2.
Now let $\mathbf{F}(\mathbf{r}) = f(\mathbf{r})\nabla g(\mathbf{r})$, with f and g scalar functions. Using that $\nabla\cdot[f(\mathbf{r})\nabla g(\mathbf{r})] = [\nabla f(\mathbf{r})]\cdot[g(\mathbf{r})] + f(\mathbf{r})\nabla^2 g(\mathbf{r})$, we obtain,

$$\int d\mathbf{r}\; f(\mathbf{r})\nabla^2 g(\mathbf{r}) \;=\; -\int d\mathbf{r}\;[\nabla f(\mathbf{r})]\cdot[g(\mathbf{r})]\;.$$

Interchanging f and g, the right hand-side does not change, so that,

$$\int d\mathbf{r}\; f(\mathbf{r})\nabla^2 g(\mathbf{r}) \;=\; \int d\mathbf{r}\; g(\mathbf{r})\nabla^2 f(\mathbf{r})\;.$$

This result represents two subsequent partial integrations, which is used in subsection 2.1 to calculate the mean-squared displacement.
Stokes's integral theorem reads, in our present notation (where, as before, $\nabla_{\hat{u}}$ the gradient operator in eq.(6), but now with differentiations with respect to the Cartesian coordinates $\hat{u}_x$, $\hat{u}_y$, and $\hat{u}_z$ of $\hat{\mathbf{u}}$),

$$\int_A d\hat{u}\;\hat{\mathbf{u}}\cdot[\,\nabla_{\hat{u}}\times\mathbf{F}(\hat{\mathbf{u}}\,] \;=\; \oint_{\partial A} d\mathbf{l}\cdot\mathbf{F}(\hat{\mathbf{u}})\;,$$

where A is a surface in three-dimensional space, and ∂A the closed boundary of A, which is a curve in three-dimensional space. We are interested here in integrations ranging over the unit-spherical surface, for which the boundary is empty. Using that $\hat{\mathbf{u}}\cdot[\nabla_{\hat{u}}\times\mathbf{F}(\hat{\mathbf{u}})] = \hat{\mathcal{R}}\cdot\mathbf{F}(\hat{\mathbf{u}})$, we thus have,

$$\oint d\hat{\mathbf{u}}\;\hat{\mathcal{R}}\cdot\mathbf{F}(\hat{\mathbf{u}}) \;=\; 0\;,$$

where the integral is understood to extend over the unit-spherical surface (which is the reason that we used $\oint$ instead of $\int$).
For a scalar function $f(\hat{\mathbf{u}})$ and a vector function $\mathbf{G}(\hat{\mathbf{u}})$, we have the identity,

$$\oint d\hat{\mathbf{u}}\; f(\hat{\mathbf{u}})\,\hat{\mathcal{R}}\cdot\mathbf{G}(\hat{\mathbf{u}}) \;=\; \oint d\hat{\mathbf{u}}\;\hat{\mathcal{R}}\cdot[\,f(\hat{\mathbf{u}})\,\mathbf{G}(\hat{\mathbf{u}})\,] - \oint d\hat{\mathbf{u}}\;\mathbf{G}(\hat{\mathbf{u}})\cdot\hat{\mathcal{R}}f(\hat{\mathbf{u}})\;. \tag{68}$$

Using the above result with $\mathbf{F} = f\,\mathbf{G}$ it is thus found that,

$$\oint d\hat{\mathbf{u}}\; f(\hat{\mathbf{u}})\,\hat{\mathcal{R}}\cdot\mathbf{G}(\hat{\mathbf{u}}) \;=\; -\oint d\hat{\mathbf{u}}\;\mathbf{G}(\hat{\mathbf{u}})\cdot\hat{\mathcal{R}}f(\hat{\mathbf{u}})\;, \tag{69}$$

which represents a partial integration that has been used in subsections 4.1 and 4.2.
Now let $\mathbf{F}(\hat{\mathbf{u}}) = f(\hat{\mathbf{u}})\hat{\mathcal{R}}g(\hat{\mathbf{u}})$, so that $\hat{\mathcal{R}}\cdot\mathbf{F}(\hat{\mathbf{u}}) = [\hat{\mathcal{R}}f(\hat{\mathbf{u}})]\cdot[\hat{\mathcal{R}}g(\hat{\mathbf{u}})] + f(\hat{\mathbf{u}})\,\hat{\mathcal{R}}^2 g(\hat{\mathbf{u}})$, and hence,

$$\oint d\hat{\mathbf{u}}\; f(\hat{\mathbf{u}})\,\hat{\mathcal{R}}^2 g(\hat{\mathbf{u}}) \;=\; -\oint d\hat{\mathbf{u}}\;[\hat{\mathcal{R}}f(\hat{\mathbf{u}})]\cdot[\hat{\mathcal{R}}g(\hat{\mathbf{u}})]\;.$$

Interchanging f and g thus shows that,

$$\oint d\hat{\mathbf{u}}\; f(\hat{\mathbf{u}})\,\hat{\mathcal{R}}^2 g(\hat{\mathbf{u}}) \;=\; \oint d\hat{\mathbf{u}}\; g(\hat{\mathbf{u}})\,\hat{\mathcal{R}}^2 f(\hat{\mathbf{u}})\;. \tag{70}$$

which represents two partial integrations. This is used in, for example, subsection 3.

References

[1] J.-P. Hansen, I.R. McDonald, *Theory of Simple Liquids*, Academic Press, 1986

[2] W.B. Russel, D.A. Saville, W.R. Schowalter, *Colloidal Dispersions*, Cambridge University Press, 1999

[3] G. Nägele, Lecture Notes 14: *The Physics of Colloidal Soft Matter*, Centre of Excellence for Advanced Materials and Structures, Warsaw, 2004

[4] J.K.G. Dhont, *An Introduction to Dynamics of Colloids*, Elsevier, 1996

[5] J. Crank, *The Mathematics of Diffusion*, Clarendon Press, 1989

[6] U. Kaatze, J. Chem. Eng. Data, 1989, **34**, 371

[7] E. Lauga, Phys. Rev. Lett. **106**, 178101 (2011)

[8] J.R. Howse, R.A.L. Jones, A.J. Ryan, T. Goughj, R. Vafabakhsh, R. Golestanian, Phys. Rev. Lett. **99**, 048102 (2007)

[9] G. Volpe, I. Buttinoni, D. Vogt, H.-J. Kümmerer, C. Bechinger, Soft Matter **7**, 8810 (2011)

[10] I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, C. Bechinger, J. Phys.:Condens. Matter **24**, 284129 (2012)

[11] J. Elgeti, R.G. Winkler, G. Gompper, to appear in Rep. Prog. Phys.

[12] G.I. Taylor, Proc. Roy. Soc. A, 1953, **219**, 186

[13] A. Vikhansky, Microfluid Nanofluid, 2009, **7**, 91

[14] D.A. McQuarrie, *Statistical Mechanics*, Harper & Row Publishers, 1973 (page 282)

[15] P. Nozieres, D. Quemada, Europhys. Lett., 1986, **2**, 129.

[16] V. Schmitt, C.M. Marques, F. Lequeux, Phys. Rev. E, 1995, **52**, 4009.

[17] J.K.G. Dhont, W.J. Briels, Rheol. Acta, 2008, **47**, 257.

[18] J.K.G. Dhont, K. Kang, M.P. Lettinga, W.J. Briels, Korea-Australia Rheology Journal, 2010, **22**, 291.

[19] J. Howon, K. Kang, K.H. Ahn, J.K.G. Dhont, Soft Matter, 2014, **10**, 9470

[20] N.F. Carnahan, K.E. Starling, J. Chem. Phys., 1969, **51**, 635.

[21] J.K.G. Dhont, Phys. Rev. E, 1999, **60**, 4534.

[22] R. Besseling, L. Isa, P. Ballesta, G. Petekidis, M. E. Cates, and W. C. K. Poon, Phys. Rev. Lett. 2010, **105**, 268301.

[23] J. Goyon, A. Colin, G. Ovarlez, A. Ajdari, L. Bocquet, Nature, 2008, **454**, 84.

[24] K.N. Pham, G. Petekidis, D. Vlassopoulos, S.U. Egelhaaf, W.C.K. Poon, P.N. Pusey, J. Rheol., 2008, **52**, 649.

[25] A. Ikeda, L. Berthier, Phys. Rev. E, 2013, **88**, 052305.

B 2 Simulation Techniques

M. Ripoll and G. A. Vliegenthart
Theoretical Soft Matter and Biophysics
Institute of Complex Systems
Forschungszentrum Jülich GmbH

Contents

1 Why do we need computer simulations? 2

2 Basic concepts 3
2.1 Interaction potentials . . . 3
2.2 System observables . . . 6

3 System size limitations 7
3.1 Periodic boundary conditions . . . 8
3.2 Potential truncation . . . 9
3.3 Neighbors list . . . 9
3.4 Long-range forces . . . 9

4 Molecular Dynamics simulations 10
4.1 Algorithms . . . 10
4.2 Adjusting the code . . . 11
4.3 Equilibration and thermostats . . . 12
4.4 Ensembles . . . 13

5 Non-equilibrium Molecular Dynamics 14

6 Monte Carlo Simulations 16

7 Summary and outlook 18

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Why do we need computer simulations?

The History of Science in general, and of Physics and Biology in particular, has fruitfully developed for centuries without the help of computer simulations until a relatively very recent date. It was in the early 1950's when the first electronic computers became available for non-military use. Before that time, the description of material properties was obtained by combining experiments and analytical approaches. The experimental observation of a system needs always of an apparatus that can be the eye, a clock, a microscope, or a nuclear magnetic resonance device. These apparatus are able to access different but limited amounts of information like average density, or the position of particle in a limited size range. To interpret and understand the experimental observations of a particular system is necessary the construction of a model, what requires the assumption of ideal conditions, and simplified interactions. The behavior of such model is normally described in terms of equations which are soluble only in a very few exceptional cases. For example, to solve the motion of more than two interacting particles, even with simple Newtonian mechanics, resulted essentially impossible before the appearance of computers. The standard procedure is then the performance of more or less dramatic approximations. Therefore, for an analytical theory to explain satisfactorily experimental results is required that, enough precision of apparatus, good choice of the system model, and validity of the performed approximations.

With the appearance of computers a new perspective can be employed to understand physical systems. The equations that describe a model with simplified interactions are solved now for a few particles or for millions of them. The algorithms necessary to perform the simulations are of course not completely problem-free, but they are in general enormously less restrictive than the analytical approximations performed in theory. Comparing results from simulations and analytical theories serves first to test the performed approximations. Comparing results from simulations and experiments serves to test the simplified interactions assumed for designing the model (see Fig. 1). Furthermore, computer simulations can be used to predict material properties, by anticipating systems or conditions that are not always easy or cheap to perform experimentally. The information provided by a simulation is in general also much more detailed than the experimental one, since very precise information of the system constituents is generated. This dual role of the simulations, that allows to bridge the distance between the models and the theory, and between the models and the experiments, makes that they are sometimes referred *theoretical simulations* by experimentalists. or as *computer experiments* by theoreticians.

Molecular Dynamics (MD) is a technique to compute steady and transport properties of many-body systems, and it was together with Monte Carlo, the first simulation method to be proposed. The basic of the approach is that the material properties are described by the Newton's equation of motion of each component. This is by considering classical interaction potentials applied to a large number of particles. Through the equation of motion, a natural time scale is built in, and the phase space is deterministically sampled. MD generates information at a microscopic level, since position and velocities of all the component atoms are known as a function of time. Macroscopic observables like pressure, or heat capacity, are obtained from the microscopic information via statistical mechanics.

Monte Carlo (MC) simulations focus in providing and efficient stochastic sampling of the configurational and conformational phase space or parts of it. The time scale is therefore not naturally built and the objective if these simulations is to obtain good approximations for statistical quantities such an expectation values, probabilities, correlation functions, or densities of states. Note that MD simulations can become extremely slow when applied to complex systems on mi-

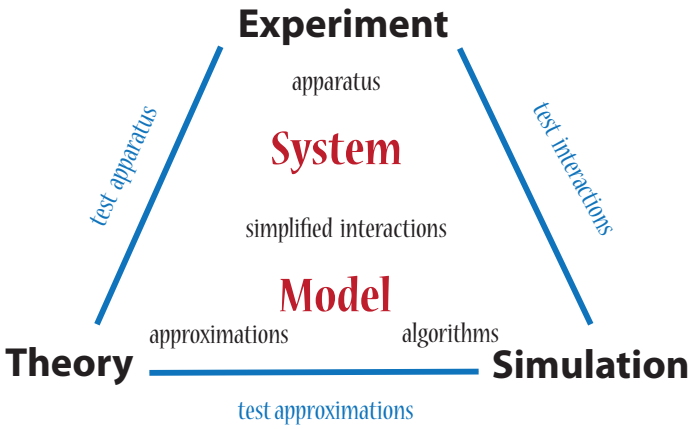


Fig. 1: *Relations between experiment, theory and simulation when analyzing a system.*

croscopic or mesoscopic scales, and that the understanding of many interesting questions does not require to consider the intrinsic dynamics of the system. In some of these systems MC can provide very effective descriptions.

Metropolis, Rosenbluth, Rosenbluth, Marshall, Teller and Teller, introduced Monte Carlo in 1953 [1] to investigate fluid properties. Alder and Wainwright introduced Molecular Dynamics in 1957 [2] to study the phase transitions of hard sphere systems, and in 1964 Rahnman carried out the first simulations using a realistic potential for liquid Argon [3]. The first simulation of a protein was done by Levitt and Warshel in 1975 [4]. Since then, the simulation techniques have substantially evolved. Particular problems have developed their own specialized techniques, like combination with quantum mechanical methods to find application at the quantum level [5], or hybrid simulation with mesoscopic techniques in order to bridge the gap with larger scales as those in soft matter systems (see Chapter B3).

In this contribution, some general concepts, common to all simulation methods will be discussed, but the main focus will be the introduction of Molecular Dynamics and Monte Carlo. For further insight in these and other simulation methods we recommend, several specialized books [6, 7, 8, 9, 10], that provide very good detailed descriptions, and 'recipes' of a large list of simulations aspects.

2 Basic concepts

2.1 Interaction potentials

The description of a system composed of N particles with positions $\mathbf{r}_i$ and $(i = 1, \ldots, N)$ can be performed in terms of the forces or interaction energies exerted on such particles. The potential energy of N interacting particles can be divided in terms of the number of particles simultaneously involved in the calculation of the interaction, this is

$$U(\mathbf{r}^N) = \sum_i u_1(\mathbf{r}_i) + \frac{1}{2}\sum_i \sum_j u_2(\mathbf{r}_i, \mathbf{r}_j) + \frac{1}{4}\sum_i \sum_j \sum_k u_3(\mathbf{r}_i, \mathbf{r}_j), \mathbf{r}_k) + \cdots \qquad (1)$$

The first contribution corresponds to the effect of external fields applied on the system, like gravity or electric fields. The second term is the pair potential and includes the force that the presence of one particle induces in a second one. The third term, report forces induced due to the presence of triples, and similarly subsequent terms can be added for increasing the number of particles in the interaction. The calculation of u_1 is then and operation of order N, and u_2 of order N^2 since all pairs of particles have to be identified. Even more costly will be to determine u_3, since there are $\mathcal{O}(N^3)$ triples on a system. Furthermore, high order contributions are typically small, and the pair potential is chosen such that gives and effective potential resulting from all the high order terms. A real two-body potential would reproduce experimental data, without and explicit dependence on system parameters like density or temperature. The normally employed effective pair potentials will though depend on system parameters.

It is in the choice of the interaction potentials where the system under study will be mostly determined. In this way, the interactions are very different when simulating an ideal gas, or a system of charged particles; when simulating flexible polymer, or a molecule with a particular defined configuration like a water molecule; or when simulating an spherical vesicle or a red blood cell. Moreover, the interaction will determine the level of description. Each simulated particle can correspond with a particular atom and its precise location inside a molecule, like DNA; or each particle can represent a group of atoms, such that the averaged properties will be to some extent similar those of the real molecule.

The interactions are divided in bonded and non-bonded. The first ones are all the intramolecular interactions, while in the later, interactions such as electrostatic or van der Waals are considered. Here some of the most commonly employed potentials are introduced (see Fig. 2).

Bond potential To model a bond between two particles, it is most common to employ the harmonic potential,

$$U^H(r_{ij}) = k^H\,(r_{ij} - b) \tag{2}$$

where b is the reference *bond length*, and k^H the spring constant. The harmonic potential is basically a Taylor approximation to more sophisticated potentials around the reference bond length.

Angle potential Between the two bonds of three consecutively linked particles an angle θ can be defined. This angle can be fixed to have a certain value θ_0 equal or different from zero,

$$U^A(\mathbf{r}^N) = k^b\,(\theta - \theta_0)\,. \tag{3}$$

The bending constant k^b determines how strongly are the deviations from the reference angle.

To fix the relative position of four consecutively linked particles can be done with torsion potentials. The dihedral angle potential, for example is used to constrain the rotation of one particular bond around the plane defined with another two bonds.

Van der Waals potential Two particles not bounded experience frequently a combination of repulsive and attractive interactions. The strong repulsive interaction at short distances are due to excluded volume effects, while the soft attraction at larger distances is a consequence of the correlation between the electron clouds surrounding the atoms ('van der Waals or 'London' dispersion). The most common potential to model the van der Waals interactions is known as

the *Lennard-Jones potential*,

$$U^{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{4}$$

where the two potential parameters are ϵ, the minimum potential depth, and σ the collision diameter.
To study systems of purely repulsive particles, a frequent approach is to consider only the repulsive part of this potential by truncating it precisely at the minimum and increasing its value to zero at that value,

$$U^{rLJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + 1 \right], \qquad r < r_c = 2^{1/6}\sigma. \tag{5}$$

This is the *repulsive Lennard Jones potential*.

Electrostatic interaction The electrostatic interaction between two particles of charge q_i, and q_j is given by the Coulomb potential,

$$U^{C}(r_{ij}) = \frac{1}{4\pi\varepsilon_0\varepsilon_R}\frac{q_i q_j}{r_{ij}} \tag{6}$$

where ε_0 is the permittivity of vacuum, and ε_R is the dielectric constant.

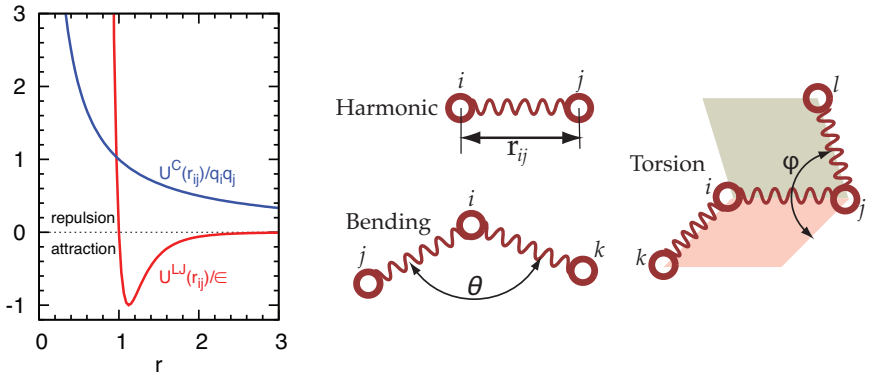


Fig. 2: *Illustration of the most basic interparticle interactions*

By combining the above potentials, the number of possible macromolecules and structures that can be modeled is unlimited. Some basic examples are shown in Fig. 3. Colloidal dispersions can be modelled by particles where the main interaction are van der Waals or Coulomb. By linking a determined number of particles with harmonic springs, the properties of polymers can be studied. If the linked particles have also a bending potential that keeps the structure elongated and stiff, we will be analyzing the properties of rod-like colloids. If the string of particles is consider simultaneously with a surrounding solvent and a few particles have an attractive interaction with the solvent and the rest a repulsive attraction with the solvent, we will have a basic lipid molecule , that with proper concentration values will be able to self-assemble.

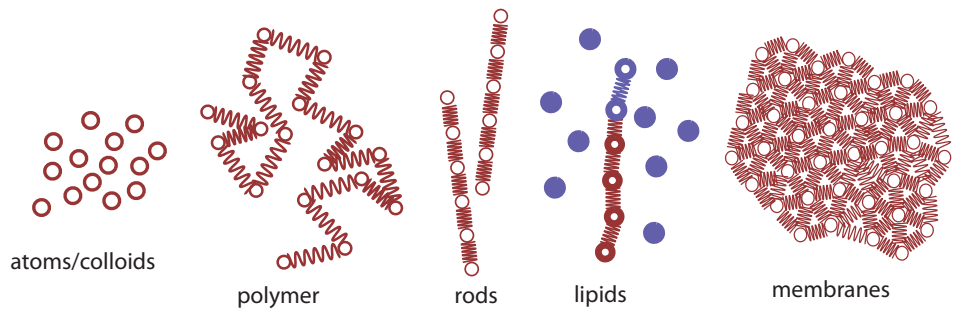


Fig. 3: *Illustration of the construction of macromolecules by combining various interaction potentials.*

2.2 System observables

The information generated by computer simulations is very detailed and lays at the microscopic level. To translate it into macroscopic terms concepts of Statistical Mechanics are to be employed [11, 12]. The number of properties that can be characterized and quantified is therefore very large. It will depend then on the particularities of the chosen system, and on the interests of the study. The system temperature T is for example defined through E_k, the system kinetic energy

$$E_K = \left\langle \frac{1}{2} \sum_i^N m_i v_i^2 \right\rangle = \frac{d}{2} N k_B T \tag{7}$$

where v_i is the particle velocity and d the system dimension. The pressure can be calculated from the virial equation,

$$P = \rho k_B T + \frac{1}{dV} \left\langle \sum_{i<j} \mathbf{F}_{ij} \cdot \mathbf{r}_{ij} \right\rangle, \tag{8}$$

where V is the system volume. Note that in the two previous expressions, averaged quantities are employed. These averages are defined when a quantity A can be independently measured N_R times. The average value of A is then,

$$\langle A \rangle = \frac{1}{N_R} \sum_{k=1}^{N_R} A_k. \tag{9}$$

In systems where time is defined, the independent measurements can be made at separated enough time points, or by repeating the simulation with different initial values. And finally, in order to estimate the accuracy of the measurement, the standard deviation of the average is calculated as

$$\delta(A) = \sqrt{\frac{1}{N_R - 1} \sum_{k=1}^{N_R} (A_k - \langle A \rangle)^2} \tag{10}$$

Apart from thermodynamic properties, static information in real or reciprocal space permits to obtain structural properties which are of fundamental when characterizing a system. The pair correlation function $g(r)$ for example, measures the probability to find a particle at a distance r from another particle. If this function shows or not some periodic behavior will allow to

determine if the system is a gas, a liquid, or a crystal (an example will be shown in Sect.4.2). Dynamical properties are also of great interest. In MD the time evolution is intrinsic such that time dependent quantities can be described. As an example, the mean squared displacement $\langle(\Delta r(t))^2\rangle = \left\langle \sum_i \left(\mathbf{r}_i(t) - \mathbf{r}_i(0)\right)^2 \right\rangle$, shows the transition from a short time ballistic regime in which $\langle(\Delta r(t))^2\rangle \sim t^2$ to a long time diffusive regime, $\langle(\Delta r(t))^2\rangle \sim t$, what allows to determine the self-diffusion coefficient D_s

$$\left\langle (\Delta r(t))^2 \right\rangle = 2dD_s t. \tag{11}$$

Reduced units Quantities like time or temperature obtained from simulation results are expressed in terms of the model parameters. In order to relate simulation results with experimental data is convenient to map the model parameters into physical units. The comparison of simulation results with the Lennard Jones potential in Eq. (4) and experimental results with Argon provides one of the most established choices for simulation units, the so called *reduced units*. These consist in choosing σ as a unit of length and relate to $\sigma = 3.40510^{-10}$ m, ϵ as a unit of energy with $\epsilon/k_B = 119.8$ K, and the reference particle mass m_i as the unit of mass with $m_i = 0.03994$ kg/mol. Some related quantities are specified in Table 1.

Quantity	Reduced unit	Simulation value	Physical value
temperature	$T^* = k_B T/\epsilon$	$T^* = 1$	$T = 119.8$ K
time	$t^* = t\sqrt{\epsilon/m_i\sigma^2}$	$t^* = 0.01$	$t = 4.32210^{-10}$ s
pressure	$P^* = P\sigma^3/\epsilon$	$P^* = 1$	$P = 41.9$ MPa
density	$\rho^* = \rho\sigma^3$	$\rho^* = 0.4$	$\rho = 960$ kg/m^3

Table 1: *Most relevant reduced units and their translation into physical units*

The main importance of the reduced units is that many other parameter combinations translate into the same reduced units, such that their results should be equivalent. This is the *law of corresponding states* .

3 System size limitations

In Statistical Mechanics, systems are frequently studied in the thermodynamic limit. The idea is that the thermodynamic variables are independent on the collectivity and the geometry of the container. This means that extensive variables are assumed to be infinitely large, while intensive variables are kept constant. For instance, density $\rho = N/V$ is supposed to be finite, although the limits of infinite number of particles $N \to \infty$, and infinite volume $V \to \infty$ are considered. The number of particles in a simulation is normally constrained by the execution time, such that 10 to 10^7 particles are typically computed. This is a very small number when compared with the Avogadro number $\mathcal{O}(10^{23})$, and also clearly too small to achieve the thermodynamic limit. Therefore, considerations about the finite size effects have always to be taken into account.
Another important aspect are the geometrical specifications for the system container. Open boundary conditions (system with no spatial limitations) can be considered only in a few exceptional cases in which the cohesive forces are large enough to keep the system confined, like in the case of a small liquid droplet, or a microcrystal. Confining geometries, like in the presence of planar walls, can also be of special interest in many systems. Nevertheless, when the interest

is to reproduce bulk properties, confining geometries have large surface effects. For example in a 3-dimensional cubic container $V = L^3$, only $N_{bulk} \sim \rho(L^3 - 6L^2)$ particles will not be directly at one of surfaces.

3.1 Periodic boundary conditions

The usual strategy to overcome surface effects is to employ periodic boundary conditions. The idea is to simulate a parallelepiped box (square or cubic) and to assume that it repeats infinitely in space, letting particles stream away through the boundaries. Due to the periodic conditions, when a particle leaves the central box, one of its images enters again the box through the opposite face (see Fig. 4), such that the number of particles in the main box is conserved, and that particles close to the boundaries are similar than those in the center of the box.

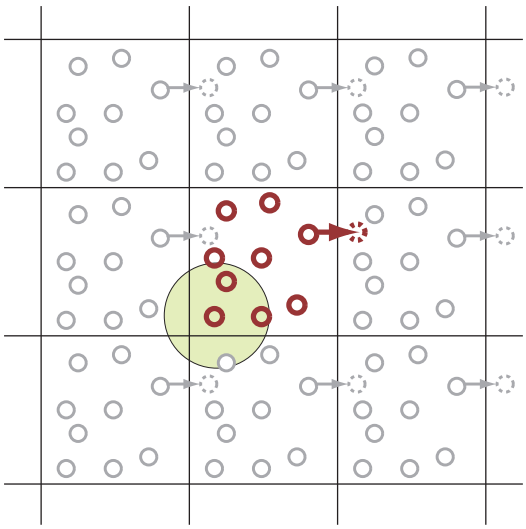


Fig. 4: *Illustration of the periodic replicas in periodic boundary conditions in two dimensions. A particle leaving the central box and entering through the opposite face is here indicated with an arrow. The radius cutoff and the minimum image convention determine the interacting neighbours of a particle close to the boundary, here indicated by the shadowed circle.*

To determine up to which extent the properties of a small and infinitely periodic system reproduces the properties of a macroscopic system depends on the phenomenon under study and the on range of the intermolecular interactions. When the potentials are long range (like Coulomb interactions) a particle will interact with its own periodic images and an anisotropic effect will be artificially introduced in the system, due to the underlying lattice structure, apart from undesired additional self-correlations. For short range potentials these effects will have much smaller impact, if the central box has a reasonably large size. For instance, in the case of a Lennard Jones fluid, a box of size $L \sim 6\sigma$ would already be reasonable. Another limitation of the use of periodic boundary conditions is the truncation of the long-wave fluctuations. Perturbations like density waves, with wave-vector larger than the box size L are simply suppressed. Simulating, for example, the properties of a liquid system close to the gas-liquid critical point is not possible within these boundaries since the range of the fluctuation is macroscopic.

Despite the previous considerations, periodic boundary conditions have extensively shown to have little effect in determining the thermodynamic properties of a large number of systems in the liquid or solid state.

3.2 Potential truncation

Most interactions considered as short range have, strictly speaking, infinitely long interaction range. As an example, the Lennard Jones potential has an attractive tail that decays with $\sim r^6$, such that the potential energy of two particle placed at a distance 3σ is practically negligible. The *potential truncation* suggests then to define a certain *cutoff radius* that introduces a distance from which the potential is approximated by zero. A typical cutoff radius for the LJ potential is $r_c = 2.5\sigma$.
When considering periodic boundary conditions, the interaction with particles in all periodic images should in principle be considered. In the case of particles with a finite interaction range, this means that a particle interacts with all those particles in the central box, and periodic images, which are placed at a distance smaller than the interaction range. This is known as the *minimum image convention* and illustrated in Fig. 4.

3.3 Neighbors list

To determine pairs of interacting particles can be a very demanding part of the code since, in principle, all particle pairs are potentially interacting. This implies an operation of order $\mathcal{O}(N^2)$. Additional tricks have been developed that reduce the cost of this operation. These tricks are only intended to improve the efficiency of the code and should not affect the system behaviour.

Verlet list Verlet [13] suggested to store a list that would indicate for each particle, at one precise time point, all the neighbouring particles within a distance $r_c + \Delta r$. Then only pair in the list are checked as interacting couples what dramatically reduces the operation. The list has to be updated when a particle has travel a distance Δr. An optimum value of Δr has to be found, since large values will include more particles in the list, but will need to be updated less frequent.

Cell list The idea of the cell list is to divide the simulation box in cells that are fixed in space, such that for each cell it can be known and stored which are the neighbouring cells. Then the particles are sorted into the cells and the interacting particles are searched only within the neighbouring cells. The cell list can be independently employed or used for instance to determine the Verlet list. This can reduce the computational time almost to $\mathcal{O}(N)$, what is very useful for very large systems.

3.4 Long-range forces

A force is considered to be long range if it does not decay faster than r^{-d}, with d the system dimension. The Coulomb interaction $\sim r^{-1}$ between charged particles, and the dipole-dipole interaction $\sim r^{-3}$ are very important examples. Naive solutions like dramatically increasing the system size are not feasible. Several specific methods to tackle the problem of the long forces have been developed. The Ewald sum method, includes the interaction of an ion with

all its periodic images. The reaction field methods assume that the interactions with molecules beyond a certain cutoff can be averaged. Details of these and further methods can be found in [6, 7, 14].

4 Molecular Dynamics simulations

The basic idea of Molecular Dynamics is that N particles of masses m_i with $(i = 1, \ldots, N)$ satisfy the Newton's equations of motion,

$$m_i \frac{d^2}{dt^2}\mathbf{r}_i = \mathbf{F}_i, \tag{12}$$

where $\mathbf{r}_i$ are the particle positions, and $\mathbf{F}_i$ are the resulting forces exerted on particle i. In the case of conservative systems, the forces are obtained from the potential energy with $\mathbf{F}_i = -\nabla U(\mathbf{r}^N)$, with $\mathbf{r}^N = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$.

4.1 Algorithms

Given a system with a large number of particles and determined interactions, the equations of motion in Eq. (12) can not be analytically solved. The approach of MD is to repeatedly evaluate all the particle positions, velocities, and forces at a certain time point, given that all the values are known at a recent time. A discretization scheme is therefore required to integrate the Newton's equation of motion. The first naive approach is the *Euler algorithm*,

$$\begin{aligned} \mathbf{r}_i(t+\delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t)\delta t + \frac{\mathbf{F}_i(t)}{2m_i}\delta t^2 \\ \mathbf{v}_i(t+\delta t) &= \mathbf{v}_i(t) + \frac{\mathbf{F}_i(t)}{m_i}\delta t. \end{aligned} \tag{13}$$

Once that the first simulation program is coded, several aspects have to be considered to decide whether the approximation achieved by the algorithm is or not acceptable. The main factors are: *(i) Energy and momentum conservation*. The problem has been stated in the microcanonical ensemble, such that energy and momentum have to be conserved. At short time there will be fluctuations around the average values, but at long times this average value has to keep constant. *(ii) Time reversibility*. Similar to the Newton equations the algorithm should be time reversible and phase space conserving. Nevertheless, the computer rounding errors make that this conservation never will be exact. *(iii) Reasonable execution times*. The length, and the number of force evaluations per time step, are essential to determine the CPU time that an algorithm requires under certain system restrictions. To find a good performing algorithm is thus of great importance.

The Euler algorithm in Eq. (13) has problems with energy conservation and with time reversibility even with very small time steps. A simple and very good algorithm is the *Verlet algorithm*. This algorithm considers a Taylor expansion of a particle position and velocity around a time t for a small time steps $\pm\delta t$,

$$\mathbf{r}_i(t \pm \delta t) = \mathbf{r}_i(t) \pm \frac{d\mathbf{r}_i(t)}{dt}\delta t + \frac{1}{2}\frac{d^2\mathbf{r}_i(t)}{dt^2}\delta t^2 \pm \frac{1}{6}\frac{d^3\mathbf{r}_i(t)}{dt^3}\delta t^3 + \mathcal{O}(\delta t^4) \tag{14}$$

summing these two expressions and neglecting terms of order δt^4,

$$\mathbf{r}_i(t+\delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t) + \frac{\mathbf{F}_i(t)}{m}\delta t^2. \tag{15}$$

This is, the position in the new time is determined by the position in the two previous time steps and the force. To determine the forces in a certain time, only the particle positions are required, since the employed forces are conservative. This implies that the velocities do not intervene in establishing the particle trajectory, although they can be calculated by subtracting the expressions in Eq. (14),

$$\mathbf{v}_i(t+\delta t) = \frac{\mathbf{r}_i(t+\delta t) - \mathbf{r}_i(t-\delta t)}{2\delta t} + \mathcal{O}(\delta t^2). \tag{16}$$

The velocities are then determined with lower precision than the position and that in order to determine $\mathbf{v}_i(t)$, $\mathbf{r}_i(t+\delta t)$ should already be known. There are numerous proposals to adapt and improve this algorithm. Some of the most frequently employed are leap-frog, or the Gear predictor corrector algorithm. The *velocity Verlet algorithm* for example provides position and velocities at the same time,

$$\begin{aligned} \mathbf{r}_i(t+\delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t)\delta t + \frac{\mathbf{F}_i(t)}{2m_i}\delta t^2 \\ \mathbf{v}_i(t+\delta t) &= \mathbf{v}_i(t) + \frac{\mathbf{F}_i(t+\delta t) + \mathbf{F}_i(t)}{2m_i}\delta t. \end{aligned} \tag{17}$$

The new forces have therefore to be calculated before the velocities can be updated.

4.2 Adjusting the code

When designing a new self-made simulation code the only reasonable option is to develop it in subsequent small steps. Starting from a very simple situation that slowly becomes more complicated, such that each aspect of the code can be independently checked. One of the most important checks is always to verify the conservation laws. Energy and momentum should be constant apart from small fluctuations around an average value. The integration algorithm and the appropriate time step have to be chosen. If the time step is too small, the running time will be unnecessary long; while if the time step is too large, the equation of motion will not be properly integrated, and this will produce problems.

As an example, we look at a system of particle interacting with the repulsive Lennard-Jones potential in Eq. (5). Initially the particles are placed in the nodes of a cubic crystal, in order to avoid particle overlap. Density and temperature parameters correspond to those of a fluid, such that the system evolves from the initial crystal into a fluid. In Fig. 5, the temperature and total energy corresponding to the system integrated with two different algorithms are displayed as a function of time. With both algorithms, there are strong changes for very short times, this is while the system is melting and reaching the equilibrium state. At longer times, it can be seen that the system integrated with the Euler algorithm keeps a drift in both contributions to the energy in spite the fact that a much smaller time step has been used. With these results, we can conclude that the velocity Verlet algorithm with $\delta t^* = 0.002$ is a reasonably good choice.

There are many other quantities that can give relevant information about our system, like pressure, velocity distribution function, or time correlations. In Fig. 6 two quantities related to the previously introduced repulsive Lennard Jones system are shown. The pair distribution

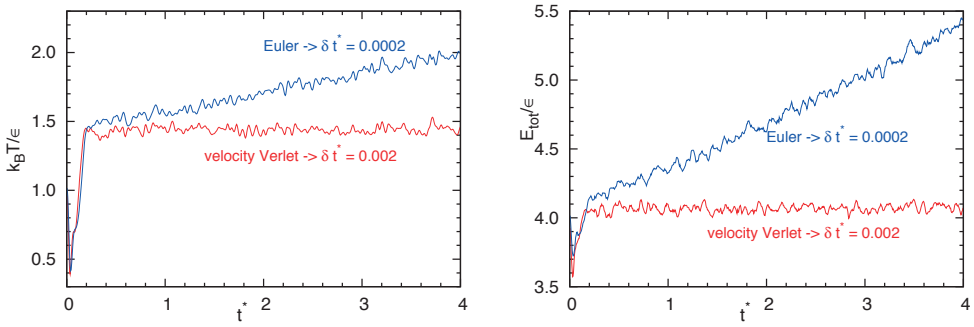


Fig. 5: *Time evolution of the averaged temperature and total energy per particle of a system of repulsive Lennard-Jones particles with $N = 10^3$, in a cubic box of size $L = 10\sigma$, and initial temperature $k_BT = \epsilon$. Results obtained with two different integration algorithms.*

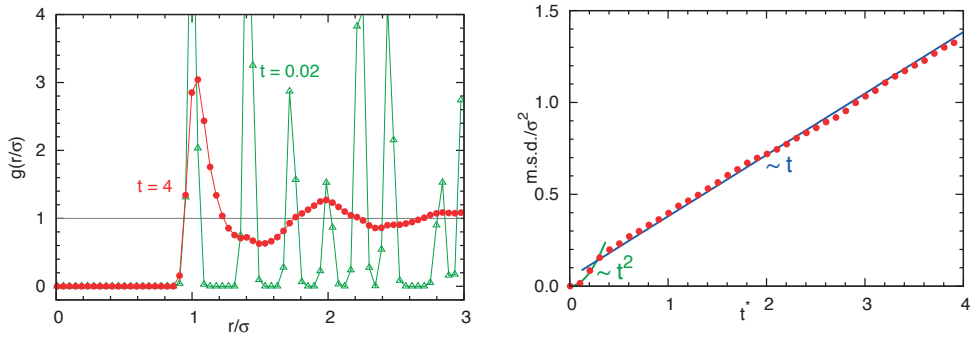


Fig. 6: *Time evolution of the mean squared displacement and pair distribution function at two times of the system in Fig. 5, integrated with the velocity Verlet algorithm.*

function $g(r)$ is computed at two different times. When the time is very short, the initial periodical structure can still be observed, while for a larger time, the characteristic distribution of a fluid is obtained. The mean squared displacement shows a ballistic short time behavior in which particles hardly interact with each other. This is given by a quadratic time dependence, $\left\langle \left(\sum_i \mathbf{r}_i(t) - \mathbf{r}_i(0)\right)^2 / N \right\rangle = dk_BT/(2m)t^2$, and it is present independently of the initially periodic configuration. The quadratic behavior shows a transition for long times into a diffusive behavior in which particles regularly collide with many others, and that it is characterized by a linear time dependence, $\left\langle \left(\sum_i \mathbf{r}_i(t) - \mathbf{r}_i(0)\right)^2 / N \right\rangle = 2dDt$, with D the self-diffusion coefficient.

4.3 Equilibration and thermostats

In many cases, as the one presented above, the system is not initialized in an equilibrium or stationary configuration. And most times these initial conditions are chosen for computational convenience and are not related to the system that we intent to study. To reach then a more relevant initial point, the simulation is let to run for certain *equilibration* time. After this time, the memory of the initial configuration is lost. The configuration after such time is then considered

as the new initial time.
It is rather common a situation like the one on Fig. 5, where the system, initially at temperature $k_BT/\epsilon = 1$, stabilizes its temperature at a different value. A usual strategy is then to rescale all the particle velocities after the equilibration time to return, or fix the temperature at the desired value with,

$$\mathbf{v}_i' = c\mathbf{v}_i. \tag{18}$$

In order to determine c, we compare the actual kinetic energy E_K with the desired temperature k_BT,

$$\frac{d}{2}Nk_BT = \frac{1}{2}\sum_i m_i(\mathbf{v}_i')^2 = c^2\frac{1}{2}\sum_i m_i\mathbf{v}_i^2 = c^2E_K \tag{19}$$

such that the scaling factor is $c = \sqrt{dNK_BT/2E_K}$. In absence of an applied flow, the initial momentum should be very precisely zero, although it usually deviates from that value during the equilibration time $\mathbf{P} = \sum_i m_i\mathbf{v}_i$, similarly to the energy E_K. The velocity rescale is now,

$$\mathbf{v}_i' = c\left(\mathbf{v}_i - \mathbf{a}\right), \tag{20}$$

with $\mathbf{a} = \mathbf{P}/\sum_i m_i$, and $c = \sqrt{dNK_BT/(2E_K - P^2/\sum_i m_i)}$.

4.4 Ensembles

The simulations introduced so far consider a system with a constant number of particles N, in a fixed volume V, at a constant temperature. Averages can be made over configurations of different realizations, or over different time points, what can be considered as an ensemble average. This means that these simulations correspond to the *microcanonical ensemble* (NVE ensemble). The characteristics of a particular problem may though require the use of a different ensemble, for which the simulations need to be adapted. The most relevant examples are following.

Canonical ensemble This is the NVT ensemble. If the system is in contact with a heat bath is the temperature and not the total energy what is a conserved quantity. There are several ways of coupling the simulation system to a thermal bath [15]. The most obvious one is by using the velocity scaling as discussed in Eq. (18) and Eq. (19), by controlling the temperature and defining new velocitites each time step. This approach does though not allow for temperature fluctuations, present in the canonical ensemble. A weaker formulation is given by the *Berenden thermostat* [16]. The velocities are scaled at each step such that the change of temperature is proportional to the temperature difference, with a coupling parameter τ that determines how tight the bath and the system are coupled together. This is extremely efficient for relaxing the system to the target temperature, but it might be important to probe a correct canonical ensemble.
The extended system method was originally introduced by Nosé [17] and subsequently developed by Hoover [18]. The idea is to consider the bath as an integral part of the system by the addition of three additional artificial variables that play the role of a time-scaling, a velocity, and a mass related quantity that determines the coupling between the reservoir and the real system and so influences the temperature fluctuations.

Isobaric-isothermal ensemble This is the NPT ensemble. In this case the volume of the system is not fixed but the pressure, the number of particles and the temperature. The idea now is that the system is couple to a pressure bath. The particle positions can be rescaled as

$$\mathbf{r}_i' = \lambda \mathbf{r}_i, \tag{21}$$

with λ calculated such that the pressure in Eq. (8) is a fixed value.

5 Non-equilibrium Molecular Dynamics

The systems under study can be made progressively more complex by increasing the number of particles involved, by distinguishing more than one type of particles, or by introducing many different interactions between particles. Extra degrees of complexity can be introduced by considering an applied external field like gravity or an electric field as stated in Eq. 1. In all these cases though the system will be studied in equilibrium conditions. In the last decades the explosion of the computer power, and the improvement of the algorithms is making accessible to study by means of simulations the properties of driven complex systems. Now I will briefly introduce how these non-equilibrium situations can be simulated with MD in some examples.

Confinement When simulating a fluid confined between walls two types of boundary conditions can be identified by characterizing the velocity of the fluid in the proximity of the walls. The fluid may 'slide' at the wall and have therefore a different velocity. This characterize the so called *slip boundary conditions*. The *stick boundary conditions* though impose that the fluid has the same velocity as the walls in their neighbourhood.
A perfectly flat wall with slip boundary conditions can be implemented by considering a repulsive Lennard Jones interaction along the wall direction. A rough wall, with boundary conditions close to stick can be included by considering a random distribution of particles with fixed position (similar to infinity mass) that interact with he fluid for example with rLJ interactions. Stick can also be obtained with bounce-back, here when a particle arrives to the wall sees its velocity inverted, such the trajectory is reverted at least in part.

Poisuelle flow If walls with stick boundary conditions are implemented together with an applied force parallel to the walls, a parabolic velocity profile will develop (see Fig. 7). These Poisuelle or capillary flows are an essential ingredient in many *microfluidic* systems where the geometry of the confinement can be made arbitrarily complex [19, 20].

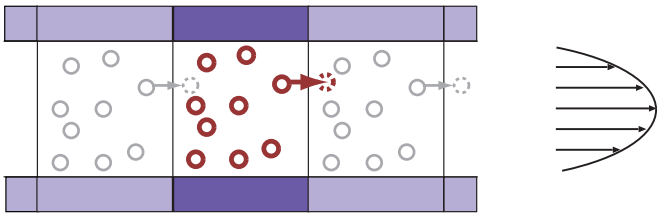


Fig. 7: *Illustration of a fluid confined between walls in one direction and periodic boundary conditions in the other. In the case of stick boundary conditions, and in the presence of an applied field, a capillary flow will be present.*

Temperature gradient Effects induced by temperature inhomogeneities are present in a large number of technical and biological systems [21, 22]. The implementation of temperature gradients is possible in combination with wall, but has been mostly done with periodic boundary conditions [23, 24]. The idea is that one cold and one hot slabs are defined in one side and in the center of the box respectively, see Fig. 8. An artificial energy transfer is then imposed from the cold to the hot slab what produces that these two slabs have indeed different average kinetic energies. The system in between the slabs will experience a physical linear temperature variation that will have opposite signs in the two simulation box halves.

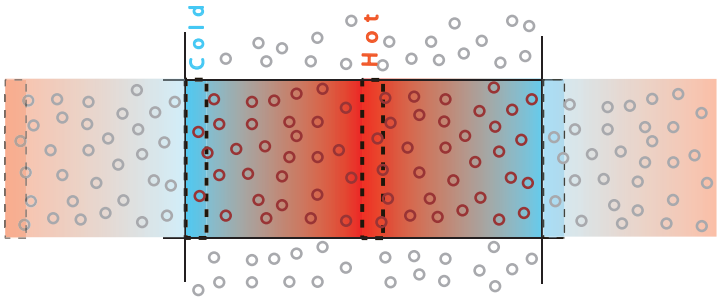


Fig. 8: *Illustration of the implementation of a temperature gradient by defining a cold and a hot slab with periodic boundary conditions.*

Shear flow When the system is in shear flow it displays a linearly varying velocity with position. This is a very interesting situation form the scientific viewpoint since it is possible to study the dynamical behaviour in a non-equilibrium steady state [19, 20]. The number of technological applications is also large ranging from food production to blood flow.

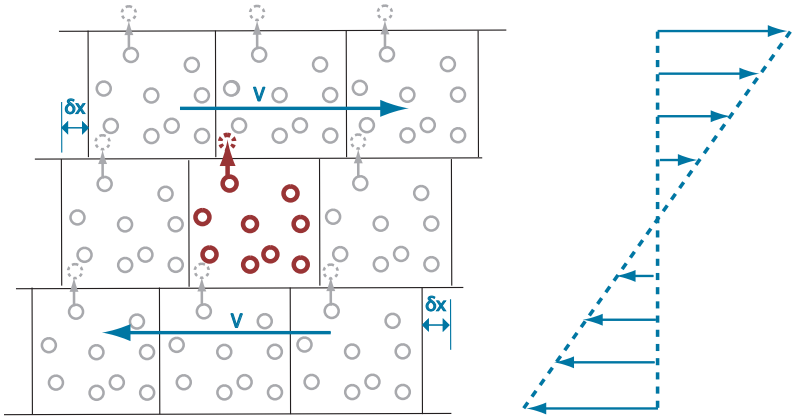


Fig. 9: *Illustration of the Lees-Edwards conditions for the implementation of a shear flow in the absence of walls.*

Shear flow can be implemented in a MD code mainly following two strategies. The first one is to confine the system between parallel walls which are moving at different velocities. The other one is a modification of the periodic boundary conditions introduced by Lees and Edwards

[25]. In essence, the infinite periodic system is subjected to a uniform shear in the flow-gradient plane. The modified boundaries are only defined in the direction of the shear flow, being the other directions standardly periodic. When a particle crosses one of the modified boundaries, it comes back through the opposite side but it will suffer an additional displacement δx in its position and a variation of its velocity u (see Fig. 9). The displacement and additional velocity transferred to the particles will be equal in magnitude but with opposite directions when crossing opposite boundaries, and their magnitudes will determine the shear rate. In this way, a shear rate $\dot{\gamma}$ is imposed in a system with L_x as the box dimension in the shear flow direction, with an additional velocity of $u = \dot{\gamma} L_x$, and a particle displacement of $\delta x = ut$, with t representing the simulation time. The average particle velocity then becomes a function of the position in the gradient direction y, $\langle v_x \rangle = \dot{\gamma} y$.

6 Monte Carlo Simulations

The main idea of the method consists in obtaining the thermodynamic properties of a system by stochastically sampling of the particle configurations proportional to their Boltzmann weight. The Boltzmann factor of a system with N particles with positions $\mathbf{r}^N$ and temperature T is

$$p(\mathbf{r}^N) = \frac{e^{-E(\mathbf{r}^N)/k_B T}}{\int d\mathbf{r}^N e^{-E(\mathbf{r}^N)/k_B T}} = \frac{e^{-E(\mathbf{r}^N)/k_B T}}{Q(N,V,T)}, \tag{22}$$

where $Q(N,V,T)$ is the partition function in the canonical ensemble. The averages of an observable $A(\mathbf{r}^N)$ which depends on the configuration $\mathbf{r}^N$ can be calculated then as,

$$\langle A \rangle = \int d\mathbf{r}^N A(\mathbf{r}^N) p(\mathbf{r}^N). \tag{23}$$

The basic idea of MC is then that instead of calculating a weighted average over all (unweighted) configurations one can also calculate an un-weighted average over weighted configurations.

Simple sampling To illustrate the Monte Carlo method we first discuss the simplest Monte Carlo approach in the calculation of integrals. This method is called random or simple Monte Carlo sampling. Suppose one has an integral of the form $I = \int_a^b f(x)dx$. Using random sampling, the value of this integral is determined by evaluating $f(x)$ at M randomly chosen values of x in the interval $\{a....b\}$ and averaging over the corresponding $f(x)$. The integral is obtained by $I = (b-a)\langle f(x) \rangle$. This method works for low dimensional integrals, and if the integrand behaves properly, i.e. is a smooth, slowly varying function of the variables.

Importance sampling However, for cases in which the integrand is a rapidly varying function of the configurations $\mathbf{r}^N$ the previous approach is very inefficient. Moreover, the integrand is zero for most configurations and sharply peaked around the average value of the energy. This behaviour of the integrand can be understood because if we were to generate configurations randomly, then most of these configurations would contain overlapping pairs of particles for which the energy is large (infinite). These configurations would have a zero Boltzmann factor. Using the simple (random sampling) Monte Carlo scheme we would thus spend a lot of time in calculating zero contributions to the integral.

For these kind of situations it is necessary to employ importance sampling. In this method, the integrand is sampled often when the Boltzmann factor is large and less frequent when the Boltzmann factor is small. Efficient schemes are then developed to generate sequences of configurations proportional to their Boltzmann factor. Using these configurations an average is calculated in which the weighing is contained in the selection of configurations $\mathbf{r}^N$. A new configuration is created from a preceding one according to a specified transition probability.

Metropolis scheme Let us now derive the Metropolis scheme [1] for determining the transition probabilities from a state $\mathbf{r}^N = o$ (o = old) to another state $\mathbf{r}'^N = n$ (n = new) and thus for generating a sequence of states obeying the Boltzmann distribution. The states o and n have Boltzmann factors given by $E(o)$ and $E(n)$ in Eq. (22). In equilibrium there is no net flow between state o and states n. This means that on average the number of accepted trial moves resulting in leaving state o must be exactly balanced by the number of accepted trial moves arriving in state o from *any other state* n (Fig. 6). Compare with for instance a gas-liquid equilibrium where *on average* per unit of time the same number of molecules leave the gaseous phase to the liquid phase, and vice versa. A stronger condition is standardly imposed, this is

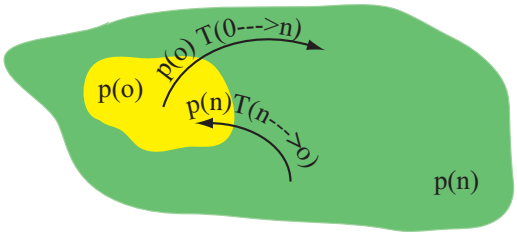


Fig. 10: *In equilibrium, the flow from a state o is balanced by the flow to state o. The Boltzmann factor of state o (n) is denoted by $p(o)$ ($p(n)$).*

that in equilibrium the number of accepted trial moves from state n to o should be exactly balanced by the number of accepted trial moves from state o to n. This condition is called detailed balance and can be written as,

$$p(o)T(o \to n) = p(n)T(n \to o) \tag{24}$$

where $T(\alpha \to \beta)$ denotes the transition probability to go from a state α to a state β. Note that in Eq. (24) the fact that the probability of finding a particular configuration is proportional to its Boltzmann factor is employed. The transition probability is in itself the product of two processes, the creation of a trial move C and the acceptance of this trial move A,

$$T(\alpha \to \beta) = C(\alpha \to \beta)A(\alpha \to \beta) \tag{25}$$

In many Monte Carlo applications the creation of a trial move is a symmetric process, i.e. the creation of the forward and the backward move have the same probability and thus $C(\alpha \to \beta) = C(\beta \to \alpha)$. Using this form, inserting Eq. (25) in Eq. (24) and using the Boltzmann weight $p(\alpha)$ in Eq. (22) gives,

$$\frac{p(n)}{p(o)} = \frac{A(o \to n)}{A(n \to o)} = \exp\left[-\left(E(n) - E(o)\right)/k_B T\right] \tag{26}$$

Following Metropolis, the final expression for the acceptance probability is,

$$\begin{aligned} A(o \to n) &= p(o)/p(n) \quad \text{if} \quad E(n) > E(o) \\ &= 1 \qquad\quad\ \ \text{if} \quad E(n) \leq E(o) \end{aligned} \tag{27}$$

In practise a Monte Carlo translational move is performed as follows. Calculate the energy of the present state $E(o)$, then select a particle at random, move this particle over a random distance $\Delta R = \mathrm{RAN} \times R_m$ with RAN a random number in the interval 0-1 and R_m the maximum displacement (which is a tunable parameter of the program). Calculate the energy of the new configuration $E(n)$. If the new energy is lower than the old energy accept the move immediately, in case the new energy is higher than the old energy then accept the move with probability $p(o)/p(n)$. This is implemented by generating a random number RAN between between 0 and 1, if $\mathrm{RAN} < p(o)/p(n)$ the move is accepted. If $\mathrm{RAN} \geq p(o)/p(n)$ the move is rejected.
The magnitude of the parameter R_m determines the efficiency of the MC procedure. If R_m is large, many of the trial moves are rejected. On the other hand, if R_m is too small, we will only sample phase space very slowly. In practise, a good choice for R_m is a value such that half of the trial moves are accepted.

Other sampling schemes Various much more sophisticated sampling methods have been and are being developed and they are employed depending on the model details and relevant questions. Some examples of these methods are Rosenbluth sampling, configurational bias sampling, parallel tempering, umbrella sampling, grand canonical MC, Gibbs ensemble MC or the Wang-Landau method.

Random Numbers By construction, random numbers play a crucial role in Monte Carlo simulations (hence the name Monte Carlo) and therefore the quality of the random number generator is of great importance. Most random number generators are algorithms that generate a sequence of numbers given an initial value of the 'seed'. The randomness of this sequence, characterised by the distribution of the numbers in the sequence, and the correlations between the numbers, determines the quality of the random number generator. A good discussion of random number generators can be found in [26], some useful algorithms are available in [27].

7 Summary and outlook

Molecular Dynamics and Mote Carlo are very versatile simulation tools, which are nowadays still very broadly employed to study the properties of a large spectrum of physical, chemical, biological, and technical systems in a varied number of conditions. The techniques has though some limitations. The most significant one is that the times and sizes accessible by simulations is limited, although it is getting progressively more realistic. With MD the system can also eventually be trapped in a particular area of the phase space, this is when typical relaxation times that would drive the system out of such state, are larger than the accessible computing times. Ergodicity would in this case not be satisfied. Another limitation is the difficulty of bridging different length and time scales that can be present in the same system, for example in system in solution where the scale of the solute and the solvent are usually order of magnitude apart. This is solved by several mesoscopic models that will be discussed in Chapter B3.

In this chapter, we have introduced the a first glimpse of the most important considerations required for an interested reader without any previous experience in simulations. Further understanding and implementation details are directed to more specialized literature.

References

[1] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.M. Teller and E. Teller, *J. Chem. Phys* **21**, 1087 (1953).

[2] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **27**, 1208 (1957).

[3] A. Rahman, Phys. Rev. **136**, A 405 (1964).

[4] M. Levitt and A. Warshel, Nature **253**, 94 (1975).

[5] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[6] M. P. Allen and D. J. Tildesley, *Computer Simulations in Liquids* (Clarendon, Oxford, 1987).

[7] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, San Diego, 2002), 2nd ed.

[8] D. J. Evans and G. P. Morriss, *Statistical Mechanics of NonEquilibrium Liquids* (Academic Press, London, 1990).

[9] W. G. Hoover, *Molecular Dynamics* (Springer, Berlin, 1990).

[10] D. W. Heermann, *Computer Simulations Methods in Theoretical Physics* (Springer, Berlin, 1990).

[11] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon Press, 1959).

[12] J. P. Hansen and I. Mc Donald, *Theory of simple liquids* (Academic, New York, 1986).

[13] L. Verlet, Phys. Rev. **159**, 98 (1967).

[14] M. Deserno and C. Holm, J. Chem. Phys. **109**, 7678 (1998).

[15] P. H. Hünenberger, Adv. Polymer Sci. **173**, 105 (2005).

[16] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

[17] S. Nosé, J. Chem. Phys. **81**, 511 (1984).

[18] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[19] R. G. Larson, *The Structure and Rheology of Complex Fluids* (Oxford University Press, New York, 1999).

[20] M. Doi and S. F. Edwards, *The Theory of Ploymer Dynamics* (Oxford University Press, Oxford, 1986).

[21] S. Wiegand, J. Phys.: Condens. Matter **16**, R357 (2004).

[22] C. J. Wienken, P. Baaske, U. Rothbauer, D. Braun, and S. Duhr, Nature Comm. **1**, 100 (2010).

[23] B. Hafskjold, T. Ikeshoji, and S. K. Ratkje, Molecular Phys. **80**, 1389 (1993).

[24] F. Müller-Plathe, J. Chem. Phys. **106**, 6082 (1997).

[25] A. W. Lees and S. F. Edwards, J. Phys. C **5**, 1921 (1972).

[26] D.P. Landau and K. Binder, *A guide to Monte Carlo simulations in statistical physics* (Cambridge University Press, Cambridge, 2000)

[27] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986)

B 3 Mesoscale Hydrodynamics

Roland G. Winkler
Institute for Advanced Simulation
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Hydrodynamics **3**
2.1 Linearised Hydrodynamics 3
2.2 Solution of the linearised Landau-Lifshitz Navier-Stokes equations 5
2.3 Velocity correlation function in Fourier space 6
2.4 Velocity correlation function in real space – long-time tail 7
2.5 Hydrodynamic tensor in real space – Oseen tensor 7

3 Mesoscale hydrodynamic simulations **8**
3.1 Lattice Boltzmann 8
3.2 Dissipative Particle Dynamics 10
3.3 Multiparticle Collision Dynamics 11

4 Hydrodynamics and Multiparticle Collision Dynamics **14**

5 Multiparticle Collision Dynamics: Transport coefficients **17**
5.1 Diffusion coefficient 17
5.2 Viscosity 18

6 MPC: Embedded object and boundary conditions **20**

7 MPC: External fields **22**
7.1 Shear flow 22
7.2 Poiseuille flow 23

8 Applications **24**
8.1 Hydrodynamic synchronization of microrotors 24
8.2 Hydrodynamics of discrete-particle models 26

9 Conclusions **28**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

During the last few decades, soft matter has developed into an interdisciplinary research field combining physics, chemistry, chemical engineering, biology, and materials science. This is driven by the specificities of soft matter, which consists of large structural units in the nano- to micrometer range and is sensitive to thermal fluctuations and weak external perturbations (cf. Fig. 1) [1–3]. Soft matter comprises traditional complex fluids such as colloidal suspensions, polymer solutions, and amphiphilic mixtures, as well as a wide range of phenomena including self-organization, transport in microfluidic devices and biological capillaries, chemically reactive flows, the fluid dynamics of self-propelled objects, and the viscoelastic behavior of networks in cells [2].

The presence of disparate time, length, and energy scales poses particular challenges for conventional computer simulation techniques. Biological systems present additional problems, because they are often far from equilibrium and are driven by strong spatially and temporally varying forces. The modeling of these systems often requires the use of coarse-grained or mesoscopic approaches that mimic the behavior of atomistic systems on the length scales of interest. The goal is to incorporate the essential features of the microscopic physics in models which are computationally efficient and are easily implemented in complex geometries and on parallel computers, and can be used to predict emergent properties, test physical theories, and provide feedback for the design and analysis of experiments and industrial applications [2]. In many situations, a simple continuum description, e.g., based on the Navier-Stokes equations is not sufficient, since molecular-level details play a central role in determining the dynamic behavior. A key issue is to resolve the interplay between thermal fluctuations, hydrodynamic interactions (HI), and spatiotemporally varying forces.

The desire to bridge the length- and time-scale gap has stimulated the development of mesoscale simulation methods (Fig. 1) such as Dissipative Particle Dynamics (DPD) [4–6], Lattice Boltzmann (LB) [7–9], Direct Simulation Monte Carlo (DSMC) [10–12], and Multiparticle Collision dynamics (MPC) [13, 14]. All the approaches are essentially alternative ways of solving the Navier-Stokes equations for the fluid dynamics. Common to them is a simplified, coarse-grained description of the fluid degrees of freedom while maintaining the essential microscopic physics on the length scales of interest.

In this contribution, several aspects of mesoscale hydrodynamics will be discussed. First of all, the basic equations of hydrodynamics, the Navier-Stokes equations, will be introduced and the relevant solutions for mesoscale systems will be discussed. This comprises the long-range character of HI and its long-time tails, aspects which have to be reproduced by mesoscale simulation approaches. Secondly, DPD, LB, and MPC will be described with a focus on MPC. The DPD method is applied in lecture F.6 for blood flow simulations. Finally, two applications of MPC will be discussed briefly, namely the synchronization of microrotors by time-dependent hydrodynamic interactions, and the hydrodynamic properties of discrete-particle colloid models.

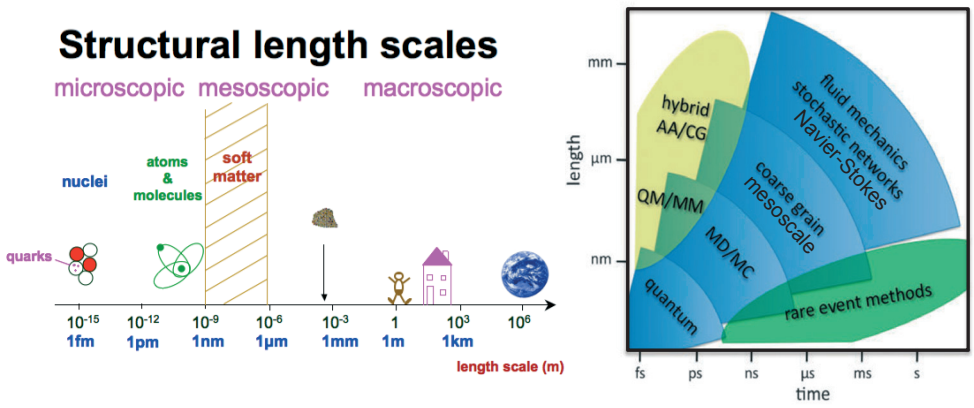


Fig. 1: *(Left) Illustration of structural lengths scales of matter. (Right) Schematics of the length and time scales of current simulation methods, ranging from atomistic quantum over classical simulation methods (molecular dynamics simulations) and mesoscale simulations to continuum methods. Advanced sampling methods, so-called rare event techniques, allow for sampling of slow processes on long-time scales. Hybrid multiscale methods provide access to large-length scales. Adopted from Ref. [15].*

2 Hydrodynamics

2.1 Linearised Hydrodynamics

Fluid flow on macroscopic scales is typically described by the Navier-Stokes equations

$$\frac{\partial}{\partial t}\varrho + \boldsymbol{\nabla}\cdot(\varrho\boldsymbol{v}) = 0, \tag{1}$$

$$\varrho\left(\frac{\partial}{\partial t}\boldsymbol{v} + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v}\right) = \boldsymbol{\nabla}\cdot\boldsymbol{\sigma} + \boldsymbol{f} \tag{2}$$

in terms of a velocity field $\boldsymbol{v}(\boldsymbol{r},t)$, where $\varrho(\boldsymbol{r},t)$ is the fluid mass density, $\boldsymbol{\sigma}(\boldsymbol{r},t)$ the stress tensor, and $\boldsymbol{f}(\boldsymbol{r},t)$ the volume force density [16–18]. Equation (1) expresses the conservation of mass. Equation (2) is an extension of Newton's equation of motion $m\dot{\boldsymbol{v}}(t) = \boldsymbol{F}$, ($m$ mass, $\boldsymbol{F}$ force) to viscous fluids and reflects the conservation of momentum. Strictly speaking there is also an equation for the conservation of energy [16–18], but we will consider isothermal systems in the following and, hence, the two equations characterize the system. The (symmetric) stress tensor of an isotropic system is given macroscopically by [16]

$$\sigma_{\alpha\beta} = -p\delta_{\alpha\beta} + \sum_{\alpha'\beta'}\eta_{\alpha\beta\alpha'\beta'}\frac{\partial v_{\alpha'}}{\partial r_{\beta'}} = -p\delta_{\alpha\beta} + \eta\left(\frac{\partial v_\alpha}{\partial r_\beta} + \frac{\partial v_\beta}{\partial r_\alpha}\right) - \delta_{\alpha\beta}\left(\frac{2}{3}\eta - \eta^{\mathrm{v}}\right)\boldsymbol{\nabla}\cdot\boldsymbol{v}, \tag{3}$$

with

$$\eta_{\alpha\beta\alpha'\beta'} = \eta(\delta_{\alpha\alpha'}\delta_{\beta\beta'} + \delta_{\alpha\beta'}\delta_{\alpha'\beta}) - \left(\frac{2}{3}\eta - \eta^{\mathrm{v}}\right)\delta_{\alpha\beta}\delta_{\alpha'\beta'}, \tag{4}$$

the local pressure $p = p(\boldsymbol{r}, t)$, the shear and bulk viscosities η and η^{v}, and α, β, α', $\beta' \in \{x, y, z\}$. Other constitutive equations can be added. We will additionally consider thermal fluctuations of the fluid, which are described by the stress tensor $\boldsymbol{\sigma}^R$, a Gaussian and Markovian stochastic process with the moments

$$\langle \boldsymbol{\sigma}^R \rangle = 0, \tag{5}$$

$$\langle \sigma^R_{\alpha\beta}(\boldsymbol{r}, t)\sigma^R_{\alpha'\beta'}(\boldsymbol{r}', t')\rangle = 2k_B T\eta_{\alpha\beta\alpha'\beta'}\delta(\boldsymbol{r} - \boldsymbol{r}')\delta(t - t'). \tag{6}$$

Here, k_B is Boltzmann's constant and T the temperature. The tensor $\boldsymbol{\eta}$ is defined in Eq. (4). Hence, $\boldsymbol{\sigma}^R$ satisfies the fluctuation-dissipation theorem [19]. With the stress tensor (3) and the random force $\boldsymbol{f}^R = \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}^R$, Eq. (2) turns into

$$\varrho\left(\frac{\partial}{\partial t}\boldsymbol{v} + (\boldsymbol{v} \cdot \boldsymbol{\nabla})\boldsymbol{v}\right) = -\boldsymbol{\nabla}p + \eta\Delta\boldsymbol{v} + \left(\frac{\eta}{3} + \eta^{\mathrm{v}}\right)\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{v}) + \boldsymbol{f} + \boldsymbol{f}^R. \tag{7}$$

In order to asses the relevance of the various terms in Eq. (7), in particular the time-dependent and non-linear inertia terms, we scale the velocity field by a typical value v_0, length by L_0, and time by T_0, as usual [20], which yields the equation

$$Re_T\,\frac{\partial \boldsymbol{v}'}{\partial t'} + Re\,(\boldsymbol{v}' \cdot \boldsymbol{\nabla}')\boldsymbol{v}' = -\boldsymbol{\nabla}'p' + \Delta'\boldsymbol{v}' + \left(\frac{1}{3} + \frac{\eta^{\mathrm{v}}}{\eta}\right)\boldsymbol{\nabla}'(\boldsymbol{\nabla}' \cdot \boldsymbol{v}') + \boldsymbol{f}' + \boldsymbol{f}^{R'}, \tag{8}$$

where the primed quantities are dimensionless and of $\mathcal{O}(1)$. Furthermore, we introduced the Reynolds numbers

$$Re = \frac{\rho v_0 L_0}{\eta} = \frac{v_0 L_0}{\nu}, \qquad Re_T = \frac{\rho L_0^2}{\eta T_0} = \frac{L_0^2}{\nu T_0}, \tag{9}$$

with the kinematic viscosity $\nu = \eta/\varrho$ [20, 21]. Typically, T_0 is defined as $T_0 = L_0/v_0$ which yields $Re_T = Re$. For a micrometer size sphere of radius R in water with thermal velocity, i.e., $R = L_0 = 10^{-3}\mu m$, $\eta = 10^{-3}Ns/m^2$, $v_0 = \sqrt{3k_BT/\varrho V} = \sqrt{9k_BT/4\varrho\pi L_0^3}$ – V is the volume –, the Reynolds number is $Re \approx 2 \times 10^{-3}$. Since the other terms are $\mathcal{O}(1)$, the left hand side of Eq. (8) is typically neglected and the Navier-Stokes equation (7) reduces to the Stokes equation. In addition, often incompressible fluids are considered, i.e., $\varrho = const.$ and $\boldsymbol{\nabla} \cdot \boldsymbol{v} = 0$, hence, Eq. (7) becomes (Stokes equation)

$$\eta\Delta\boldsymbol{v} - \boldsymbol{\nabla}p + \boldsymbol{f} = 0 \tag{10}$$

without thermal fluctuations [16–18, 20, 22]. In particular, $Re = 0$ is assumed for microswimmers, which leads to peculiarities in their locomotion as expressed by the scallop theorem [23]. The oscillatory Reynolds number Re_T can be written as $Re_T = \tau_\nu/T_0$, with $\tau_\nu = L_0^2/\nu$. Hence, Re_T is the ratio of the viscose time scale τ_ν for shear wave propagation over the distance L and the characteristic system time T_0. In order to establish proper hydrodynamic interactions, $\tau_\nu/T_0 < 1$ and, hence, $Re_T < 1$.

The relevance of the unsteady acceleration term with the oscillatory Reynolds number Re_T depends on the time scale of the physical phenomenon of interest. We will keep this term to analyze time correlations in fluids. However, we will only consider linearized hydrodynamics, which means only small deviations from stationary values. With a zero stationary velocity and

$\varrho = \rho + \delta\varrho$, we obtain the linearised Landau-Lifshitz Navier-Stokes equations

$$\frac{\partial}{\partial t}\delta\varrho + \rho\boldsymbol{\nabla}\cdot\boldsymbol{v} = 0, \tag{11}$$

$$\rho\frac{\partial}{\partial t}\boldsymbol{v} = -\boldsymbol{\nabla}p + \eta\Delta\boldsymbol{v} + \left(\frac{\eta}{3} + \eta^{\mathrm{v}}\right)\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{v}) + \boldsymbol{f} + \boldsymbol{f}^R. \tag{12}$$

These are two equations for the two unknown quantities p and $\boldsymbol{v}$. Taking the divergence of Eq. (12), we arrive at the equation

$$\Delta p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \nabla\cdot\left(\eta\Delta\boldsymbol{v} + \left(\frac{\eta}{3} + \eta^{\mathrm{v}}\right)\nabla(\nabla\cdot\boldsymbol{v}) + \boldsymbol{f}^R\right). \tag{13}$$

The second derivative with respect to time on the left-hand side follows from Eq. (11) together with the ideal gas equation of state; c is the isothermal velocity of sound.

2.2 Solution of the linearised Landau-Lifshitz Navier-Stokes equations

In order to solve the linear equations (12) and (13), we apply the Fourier transformation

$$\boldsymbol{v}(\boldsymbol{r},t) = \frac{1}{2\pi}\sum_{\boldsymbol{k}}\int\hat{\boldsymbol{v}}(\boldsymbol{k},\omega)e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}e^{i\omega t}d\omega, \tag{14}$$

$$\hat{\boldsymbol{v}}(\boldsymbol{k},\omega) = \frac{1}{V}\int\boldsymbol{v}(\boldsymbol{r},t)e^{i\boldsymbol{k}\cdot\boldsymbol{r}}e^{-i\omega t}d^3r dt \tag{15}$$

for the velocity field, with $k_\alpha = 2\pi n_\alpha/L$, $n_\alpha \in \mathbb{Z}$, and $\boldsymbol{k} \neq 0$. We use periodic boundary conditions for the spatial coordinates, since the results will later be compared with simulations. This yields ($\boldsymbol{f} = 0$)

$$\begin{aligned} i\omega\rho\hat{\boldsymbol{v}} &= i\boldsymbol{k}\hat{p} - \eta k^2\hat{\boldsymbol{v}} - \left(\frac{\eta}{3} + \eta^{\mathrm{v}}\right)k^2\mathbf{P}\hat{\boldsymbol{v}} + \hat{\boldsymbol{f}}^R, \\ \left(\frac{\omega^2}{c^2} - k^2\right)\hat{p} &= i\boldsymbol{k}\cdot\left(\eta k^2\hat{\boldsymbol{v}} + \left(\frac{\eta}{3} + \eta^{\mathrm{v}}\right)k^2\mathbf{P}\hat{\boldsymbol{v}} - \hat{\boldsymbol{f}}^R\right), \end{aligned} \tag{16}$$

where $\mathbf{P}$ is a projection operator with the components $P_{\alpha\beta} = k_\alpha k_\beta/k^2$, which projects a vector along the direction of $\boldsymbol{k}$, and $k = |\boldsymbol{k}|$. With the separation $\hat{\boldsymbol{v}} = \hat{\boldsymbol{v}}^L + \hat{\boldsymbol{v}}^T$ into a longitudinal $\hat{\boldsymbol{v}}^L$ and transverse part $\hat{\boldsymbol{v}}^T$ with respect to $\boldsymbol{k}$, i.e., $\hat{\boldsymbol{v}}\cdot\boldsymbol{k} = \hat{v}^L k$ and $\hat{\boldsymbol{v}}^T\cdot\boldsymbol{k} = 0$, Eqs. (16) yield

$$\hat{\boldsymbol{v}}(\boldsymbol{k},\omega) = \left(\hat{\mathbf{Q}}^L + \hat{\mathbf{Q}}^T\right)\hat{\boldsymbol{f}}^R, \tag{17}$$

with

$$\hat{\mathbf{Q}}^L = \left(\tilde{\eta}k^2 + \frac{i\rho}{\omega}[\omega^2 - c^2k^2]\right)^{-1}\mathbf{P} = \hat{Q}^L\mathbf{P}, \tag{18}$$

$$\hat{\mathbf{Q}}^T = \left(\eta k^2 + i\rho\omega\right)^{-1}(\mathbf{E} - \mathbf{P}) = \hat{Q}^T(\mathbf{E} - \mathbf{P}), \tag{19}$$

and $\tilde{\eta} = 4\eta/3 + \eta^{\mathrm{v}}$; $\mathbf{E}$ is the unit matrix [24].

2.3 Velocity correlation function in Fourier space

With the help of the correlation function for the random force

$$\left\langle \hat{f}^R_\alpha(\boldsymbol{k},\omega)\hat{f}^R_\beta(\boldsymbol{k}',\omega')\right\rangle = -\sum_{\alpha',\beta'} k_{\alpha'}k'_{\beta'}\left\langle \hat{\sigma}^R_{\alpha\alpha'}(\boldsymbol{k},\omega)\hat{\sigma}^R_{\beta\beta'}(\boldsymbol{k}',\omega')\right\rangle$$
$$= \frac{4\pi k_BT}{V}\boldsymbol{k}^2\left[\eta\delta_{\alpha\beta}+\left(\frac{1}{3}\eta+\eta^{\mathrm{v}}\right)\frac{k_\alpha k_\beta}{k^2}\right]\delta(\omega+\omega')\delta_{\boldsymbol{k},-\boldsymbol{k}'}, \qquad (20)$$

we can easily calculate velocity autocorrelation functions in Fourier space. At first, we find

$$\langle\hat{\boldsymbol{v}}(\boldsymbol{k},\omega)\cdot\hat{\boldsymbol{v}}(\boldsymbol{k}',\omega')\rangle = \frac{4\pi k_BT}{V}k^2\left(2\eta|\hat{Q}^T|^2+\tilde{\eta}|\hat{Q}^L|^2\right)\delta(\omega+\omega')\delta_{\boldsymbol{k},-\boldsymbol{k}'} \qquad (21)$$

via Eqs. (17), (18), and (19). The factor 2 in front of $|Q^T|^2$ reflects the two transverse components of vorticity, and $\tilde{\eta}=4\eta/3+\eta^{\mathrm{v}}$.
The time-dependent correlation function $\langle\boldsymbol{v}(\boldsymbol{k},t)\cdot\boldsymbol{v}(\boldsymbol{k}',0)\rangle$ follows by convolution

$$\langle\boldsymbol{v}(\boldsymbol{k},t)\cdot\boldsymbol{v}(\boldsymbol{k}',0)\rangle = \frac{2k_BTk^2}{V}\delta_{\boldsymbol{k},-\boldsymbol{k}'}\int\left[2\eta Q^T(\boldsymbol{k},t-t')Q^T(\boldsymbol{k}',-t')+\tilde{\eta}Q^L(\boldsymbol{k},t-t')Q^L(\boldsymbol{k}',-t')\right]dt'. \qquad (22)$$

Fourier transformation yields:
(i) Transverse hydrodynamic function

$$Q^T(\boldsymbol{k},t)=\frac{1}{\rho}e^{-\nu k^2t}\Theta(t). \qquad (23)$$

Here, $\Theta(t)$ is Heaviside's function.
(ii) Longitudinal hydrodynamic function

$$Q^L(\boldsymbol{k},t)=\frac{1}{\rho}e^{-k^2\tilde{\nu}t/2}\left[\cos(\Omega t)-\sqrt{\frac{k^2\tilde{\nu}^2}{4c^2-k^2\tilde{\nu}^2}}\sin(\Omega t)\right]\Theta(t) \qquad (24)$$

for $4c^2/(k^2\tilde{\nu}^2)>1$ and with the abbreviation $\Omega=k^2\tilde{\nu}\sqrt{4c^2/(k^2\tilde{\nu}^2)-1}/2$. The expression for $4c^2/(k^2\tilde{\nu}^2)<1$ follows by analytical continuation [24].

Hence, evaluation of the integral in Eq. (22) yields:

(i) Transverse velocity autocorrelation function

$$\left\langle\boldsymbol{v}^T(\boldsymbol{k},t)\cdot\boldsymbol{v}^T(-\boldsymbol{k},0)\right\rangle=\frac{2k_BT}{\rho V}e^{-\nu k^2|t|}. \qquad (25)$$

(ii) Longitudinal velocity autocorrelation function

$$\left\langle v^L(\boldsymbol{k},t)v^L(-\boldsymbol{k},0)\right\rangle=\frac{k_BT}{\rho V}e^{-\tilde{\nu}k^2|t|/2}\left[\cos(\Omega|t|)-\sqrt{\frac{k^2\tilde{\nu}^2}{4c^2-k^2\tilde{\nu}^2}}\sin(\Omega|t|)\right]. \qquad (26)$$

The transverse correlation function decays simply in an exponential manner, with a characteristic time $\tau_\nu=1/\nu k^2$ determined by the kinematic viscosity and the wave vector. The longitudinal correlation function decays with a different factor $\tilde{\nu}/2$ and exhibits oscillations with the frequency Ω. Examples are provided in Fig. 5.

2.4 Velocity correlation function in real space – long-time tail

Adopting the Lagrangian description of the fluid, where a fluid element is followed as it moves through space and time, we additionally average the correlation function over the distribution of displacements $\boldsymbol{r} - \boldsymbol{r}'$. Hence, the Fourier transformation of the relation (22) leads to

$$\langle \boldsymbol{v}(t) \cdot \boldsymbol{v}(0) \rangle = \sum_{\boldsymbol{k}} \langle \boldsymbol{v}(\boldsymbol{k}, t) \cdot \boldsymbol{v}(-\boldsymbol{k}, 0) \rangle \left\langle e^{-i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} \right\rangle . \tag{27}$$

Assuming a diffusive motion of the fluid element, with Gaussian distributed displacements, we find

$$\begin{aligned} \langle \boldsymbol{v}(t) \cdot \boldsymbol{v}(0) \rangle &= \sum_{\boldsymbol{k}} \langle \boldsymbol{v}(\boldsymbol{k}, t) \cdot \boldsymbol{v}(-\boldsymbol{k}, 0) \rangle \exp\left(-k^2 \left\langle (\boldsymbol{r}(t) - \boldsymbol{r}(0))^2 \right\rangle /6\right) \\ &= \frac{k_B T}{V} \sum_{\boldsymbol{k}} \left[2Q^T(\boldsymbol{k}, t) + Q^L(\boldsymbol{k}, t) \right] e^{-k^2 D t}. \end{aligned} \tag{28}$$

Here, $\langle (\boldsymbol{r}(t) - \boldsymbol{r}(0))^2 \rangle$ is the mean square displacement, which, in the simplest case, reduces to $\langle (\boldsymbol{r}(t) - \boldsymbol{r}(0))^2 \rangle = 6Dt$, with D the diffusion coefficient.
In general, the sum over $\boldsymbol{k}$ in Eq. (28) cannot be evaluated analytically. For the transverse velocity correlation function, however, we obtain the expression

$$\left\langle \boldsymbol{v}^T(t) \cdot \boldsymbol{v}^T(0) \right\rangle = \frac{2k_B T}{\rho (2\pi)^3} \int e^{-\nu k^2 t} e^{-D k^2 t} d^3k = \frac{k_B T}{4\rho} \frac{1}{[\pi(\nu + D)t]^{3/2}}. \tag{29}$$

in the limit of an infinitely large system ($L \to \infty$, $\sum_{\boldsymbol{k}} \to V/(2\pi)^3 \int d^3k$). Hence, we find the well-known long-time tail of the transverse velocity correlation function – a major characteristics of hydrodynamics [24–33].

2.5 Hydrodynamic tensor in real space – Oseen tensor

We will now consider the mean velocity field $\boldsymbol{u}(\boldsymbol{r}, t) = \langle \boldsymbol{v}(\boldsymbol{r}, t) \rangle$, where the thermal fluctuations $\boldsymbol{f}^R$ have been averaged out. Focusing on the transverse part of the hydrodynamic tensor only, e.g., for an incompressible fluid, the flow field can be expresses as

$$\boldsymbol{u}(\boldsymbol{r}, t) = \sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \int_0^t \mathbf{Q}^T(\boldsymbol{k}, t - t') \boldsymbol{f}(t')\, dt' \tag{30}$$

starting from Eq. (17). We will assume that the force $\boldsymbol{f}$ changes much more slowly with time than the tensor $\mathbf{Q}^T$. Then, $\boldsymbol{f}$ can be taken out of the integral and the upper integration limit can be extended to infinity. Hence,

$$\boldsymbol{u}(\boldsymbol{r}, t) = \sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \frac{1}{\eta k^2} (\mathbf{E} - \mathbf{P})\, \boldsymbol{f}(t) = \sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \mathbf{Q}^T(\boldsymbol{k}) \boldsymbol{f}(t). \tag{31}$$

For an infinite system, Fourier transformation yields $\boldsymbol{u}(\boldsymbol{r}, t) = \mathbf{Q}^T(\boldsymbol{r}) \boldsymbol{F}$, with $\boldsymbol{F} = V\boldsymbol{f}$ and with $\mathbf{Q}^T(\boldsymbol{k})$, $\mathbf{Q}^T(\boldsymbol{r})$ the well-known Oseen tensors in Fourier and real space, respectively [20, 22], where

$$\mathbf{Q}^T(\boldsymbol{k}) = \frac{1}{\eta k^2} (\mathbf{E} - \mathbf{P}) \ , \quad \mathbf{Q}^T(\boldsymbol{r}) = \frac{1}{8\pi\eta|\boldsymbol{r}|} \left(\mathbf{E} + \frac{\boldsymbol{r}\boldsymbol{r}^T}{|\boldsymbol{r}|^2} \right). \tag{32}$$

Here, $\boldsymbol{r}^T$ is the transpose of the vector $\boldsymbol{r}$. The Oseen tensor reflects another important property of hydrodynamic interactions, namely their long-range character – $Q(r) \sim 1/r$. Note that the same tensor follows directly as solution of the Stokes equation (10) for an incompressible fluid.

The various hydrodynamic properties have to been accounted for by hydrodynamic simulation algorithms.

3 Mesoscale hydrodynamic simulations

The macroscopic flow behavior is rather similar for many fluids even though the microscopic structure is quite different. The microscopic details may influence the absolute values of transport coefficients, however, the form of the macroscopic hydrodynamic equations is solely determined by symmetries and conservation laws. This is the basis for various mesoscale hydrodynamic simulations approaches, which select particular "microscopic" dynamical equations, but yield hydrodynamics on large scales and long times.

3.1 Lattice Boltzmann

Boltzmann equation

The starting point of LB is the Boltzmann equation, a partial differential equation for the probability density function $f(\boldsymbol{r}, \boldsymbol{v}, t)$ in the single-particle phase space $(\boldsymbol{r}, \boldsymbol{v})$ [34–36]. The time evolution of f is governed by the equation

$$\left(\frac{\partial}{\partial t} + \boldsymbol{v}\cdot\boldsymbol{\nabla} + \frac{\boldsymbol{F}}{m}\boldsymbol{\nabla}_v\right) f(\boldsymbol{r}, \boldsymbol{v}, t) = (\partial f)_{\text{coll}}. \tag{33}$$

Here, m is the mass of the particle and $\boldsymbol{F}$ an external force. The left-hand side describes propagation (streaming) along the particle trajectory. The term $(\partial f)_{\text{coll}}$ accounts for interactions with other particles (collisions). This term is rather complex, but a closed equation can be obtained by considering point particles and binary interactions only, and, in addition apply the molecular chaos assumption, i.e., assume that the velocities and positions are independent (Stoßzahlansatz) [34]. Here, we will consider a further simplification by considering a force-free system, i.e., $\boldsymbol{F} = 0$. A system initially out of equilibrium will then relax back to the equilibrium distribution function f^{eq}. Hence, the right-hand side of Eq. (33) is replaced by $(\partial f)_{\text{coll}} = -(f - f^{\text{eq}})/\tau_r$ – the so-called BGK (Bhatnagra, Gross, and Krook) approximation [37]–, and the Boltzmann equation becomes

$$\left(\frac{\partial}{\partial t} + \boldsymbol{v}\cdot\boldsymbol{\nabla}\right) f(\boldsymbol{r}, \boldsymbol{v}, t) = -\frac{1}{\tau_r}\left(f(\boldsymbol{r}, \boldsymbol{v}, t) - f^{\text{eq}}\right), \tag{34}$$

with the characteristic relaxation time τ_r.

Lattice Boltzmann: Discretization of the Boltzmann equation

In the LB method, equation (34) is discretized with velocities $\boldsymbol{c}_k$ propagating in particular directions only. Technically, this implies a dramatic reduction of the degrees of freedom associated with the velocity space. This enhances the performance in a simulation, but measures have to

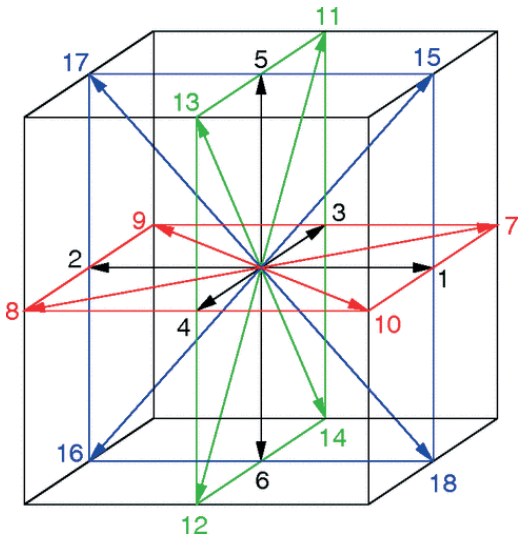


Fig. 2: *Lattice Boltzmann method: Illustration of the D3Q19 model. There are 19 populations on every lattice site: six that move to the nearest neighbors during streaming (1-6), twelve that move to to the next nearest neighbors (7-18), and the central population stays on the same lattice site.*

be taken to ensure a sufficient accurate reproduction of the flow field. The discrete Boltzmann equation is given by

$$f_k(\boldsymbol{r} + h\boldsymbol{c}_k, t + h) = f_k(\boldsymbol{r}, t) - \frac{h}{\tau_r}\left(f_k(\boldsymbol{r}, t) - f_k^{\mathrm{eq}}(\boldsymbol{r})\right) \tag{35}$$

along the direction $\boldsymbol{c}_k$. The two sides of the equation can be interpreted as follows. The right-hand side is an instantaneous local process, which changes $f_k(\boldsymbol{r}, t)$ into a $f_k^*(\boldsymbol{r}, t)$ according to the operator $\left[f_k(\boldsymbol{r}, t) - f_k^{\mathrm{eq}}(\boldsymbol{r})\right] h/\tau_r$, i.e., $f_k^*(\boldsymbol{r}, t) = f_k(\boldsymbol{r}, t) - \left[f_k(\boldsymbol{r}, t) - f_k^{\mathrm{eq}}(\boldsymbol{r})\right] h/\tau_r$. This process occurs due to collisions at the local level and is therefore called collision step. The second part of the Lattice Boltzmann equation is the streaming step which assigns the post-collisional population $f_k^*(\boldsymbol{r}, t)$ at $\boldsymbol{r}$ and t to the new population at $\boldsymbol{r} + h\boldsymbol{c}_k$ and $t + h$ [36]. The order can be exchanged and has no effect on the stationary state.
Highly symmetric lattices are used in LB, which are commonly denoted as $DnQm$ according to their dimension n (1 for 1D, 2 for 2D and 3 for 3D) and m refers to the velocity model, i.e., the number of linkages by the vectors $\boldsymbol{c}_k$. An example for the 3D lattice $D3Q19$ with 19 linkages is shown in Fig. 2. The most frequently used lattices are $D1Q3$, $D2Q9$, $D3Q15$, or $D3Q19$.

Equilibrium distribution function and conserved quantities

In order to achieve proper hydrodynamics, the following conservation laws have to be full-filled:

- Mass conservation

$$\varrho = m \int f d^3v = m \sum_k f_k = m \sum_k f_k^{\mathrm{eq}} \tag{36}$$

- Momentum conservation

$$\varrho \boldsymbol{u} = m \int \boldsymbol{v} f d^3v = m \sum_k \boldsymbol{c}_k f_k = m \sum_k \boldsymbol{c}_k f_k^{\mathrm{eq}} \tag{37}$$

- Trace of stress tensor

$$-\boldsymbol{\sigma} = m \sum_k \boldsymbol{c}_k \boldsymbol{c}_k^T f_k^{\mathrm{eq}}. \tag{38}$$

Only the trace of the stress tensor is conserved, which corresponds to the fluid energy, or, ideal gas pressure.
The discrete local equilibria f_k^{eq} are typically given by a second-order expansion in the Mach-number ($Ma = u/c$) of a local Maxwellian

$$f_k^{\mathrm{eq}}(\boldsymbol{r}, t) = \frac{w_k \varrho}{m} \left(1 + A \boldsymbol{u} \cdot \boldsymbol{c}_k + B(\boldsymbol{u} \cdot \boldsymbol{c}_k)^2 + C \boldsymbol{u}^2\right), \tag{39}$$

where w_k is a set of weights normalized to unity. Equations (36) – (38) yield the coefficients $A = 1/c^2$, $B = 1/2c^4$, $C = -1/2c^2$, where $c^2 = \sum_k w_k \boldsymbol{c}_k^2$ is the lattice sound speed, and $w_k = 1/3$ for $\boldsymbol{c}_k = 0$, $w_k = 1/18$ for $\boldsymbol{c}_k^2 = a^2/\tau^2$ (nearest neighbors 1-6), and $w_k = 1/36$ for $\boldsymbol{c}_k^2 = 2a^2/\tau^2$ (next-nearest neighbors 7-18); a is the lattice spacing. The sound velocity turns into $c = \sqrt{k_B T/m} = a/\sqrt{3}\tau$ [36].
The relaxation time τ determines the viscosity of the LB fluid. It is linked with the kinematic viscosity ν according to

$$\nu = c^2 \left(\tau - \frac{1}{2} h\right) \tag{40}$$

and follows via a Chapman-Enskog expansion [38].

3.2 Dissipative Particle Dynamics

In DPD a fluid is represented by N point particles of mass m_i, position $\boldsymbol{r}_i$, and velocity $\boldsymbol{v}_i$ ($i \in \{1, \ldots, N\}$). The particles may be considered as "fluid clusters" rather than individual atoms or molecules. Their dynamics is governed by Newton's equation of motion

$$m_i \frac{d^2}{dt^2} \boldsymbol{r}_i = \sum_{j \neq i}^{N} \left(\boldsymbol{F}_{ij}^C + \boldsymbol{F}_{ij}^D + \boldsymbol{F}_{ij}^R\right) \tag{41}$$

and they interact through pairwise forces, namely conservative $\boldsymbol{F}_{ij}^C$, dissipative $\boldsymbol{F}_{ij}^D$, and random forces $\boldsymbol{F}_{ij}^R$ [4–6, 39–41]. The main difference to a "standard" Langevin equation is the pairwise nature of the forces. It is this aspect, which ensures linear and angular momentum conservation and hence the emergence of hydrodynamics. Explicitly, the forces are given by

$$\boldsymbol{F}_{ij}^C = F_{ij}^C \boldsymbol{e}_{ij}, \tag{42}$$

$$\boldsymbol{F}_{ij}^D = -\gamma w^D(r_{ij}) \left(\boldsymbol{v}_{ij} \cdot \boldsymbol{e}_{ij}\right) \boldsymbol{e}_{ij}, \tag{43}$$

$$\boldsymbol{F}_{ij}^R = \sigma w^R(r_{ij}) \xi_{ij} \boldsymbol{e}_{ij}, \tag{44}$$

where $\boldsymbol{e}_{ij} = \boldsymbol{r}_{ij}/r_{ij}$, $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, and $\boldsymbol{v}_{ij} = \boldsymbol{v}_i - \boldsymbol{v}_j$. w^D and w^R are weight functions; γ and σ determine the amplitude of the dissipative and random forces. The $\xi_{ij} = \xi_{ji}$ are normally distributed random variables with zero mean and unite variance [40]. To satisfy detailed balance, which is sufficient to guaranteeing that the system has a Gibbsian equilibrium, requires ($m_i = m_j = m, \forall\, i.j$) [6]

$$w^D(r) = \left(w^R(r)\right)^2 \quad \text{and} \quad \sigma^2 = 2\gamma k_B T/m. \tag{45}$$

This condition renders DPD a thermostat, which is Galilean invariant and preserves hydrodynamic interactions [40, 42]. The weight function $w^R(r)$ is chosen as

$$w^R(r_{ij}) = \begin{cases} \left(1 - \frac{r_{ij}}{r_c}\right)^p & \text{for} \quad r_{ij} \le r_c, \\ 0 & \text{for} \quad r_{ij} > r_c, \end{cases} \tag{46}$$

where $p = 1$ is the standard value. However, other choices have been proposed to increase the Schmidt number, e.g., $p = 1/4$ [43]. The weight of the conservative force is typically set to

$$F_{ij}^C(r_{ij}) = \begin{cases} a_{ij}\left(1 - \frac{r_{ij}}{r_c}\right) & \text{for} \quad r_{ij} \le r_c, \\ 0 & \text{for} \quad r_{ij} > r_c, \end{cases} \tag{47}$$

with a_{ij} the conservative force coefficient between particle i and j, and the cutoff radius r_c. The equations of motion (41) are usually integrated with a modified velocity-Verlet algorithm [44], which is rather efficient in terms of both, computational cost and degree of accuracy [40]. In the review article [40], further aspects of DPD are discussed.

3.3 Multiparticle Collision Dynamics

In MPC, the fluid is represented by N point-like particles of mass m. The algorithm consists of individual streaming and collision steps (cf. Fig. 3). In the streaming step, the particles move independent of each other and experience only possibly present external forces. Without such forces, they move ballistically and their positions $\boldsymbol{r}_i$ are updated according to

$$\boldsymbol{r}_i(t + h) = \boldsymbol{r}_i(t) + h\boldsymbol{v}_i(t), \tag{48}$$

where $\boldsymbol{v}_i$ is the velocity of particle i, and h is the time interval between collisions, which will be denoted as collision time. In the collision step, a coarse-grained interaction between the fluid particles is imposed by a stochastic process. For this purpose, the system is divided in cubic cells of side length a to define the collisional environment. An elementary requirement is that the stochastic process conserves momentum on the collision-cell level, only then hydrodynamic interactions are present in the system. There are various possibilities for such a process.

(i) Stochastic rotation dynamics (SRD)

Originally, the rotation of the relative velocities $\boldsymbol{v}_i - \boldsymbol{v}_{cm}$, with respect to the center-of-mass velocity $\boldsymbol{v}_{cm}$ of the cell, around a randomly orientated axis by a fixed angle α has been suggested [2, 13, 14, 45], i.e,

$$\boldsymbol{v}_i(t + h) = \boldsymbol{v}_i(t) + (\mathbf{R}(\alpha) - \mathbf{E})\left(\boldsymbol{v}_i(t) - (t)\right), \tag{49}$$

where $\mathbf{R}$ is the rotation matrix (cf. Appendix) and

$$\boldsymbol{v}_{cm} = \frac{1}{N_c} \sum_{i=1}^{N_c} \boldsymbol{v}_i \tag{50}$$

of the N_c particles contained in the cell of particle i. The orientation of the rotation axis is chosen randomly for every collision cell and time step. As is easily shown, the algorithm conserves mass, momentum, and energy in every collision cell, which leads to long-range correlations between particles. The velocity distribution is given by the Maxwell-Boltzmann distribution in the limit $N \to \infty$.

(ii) Gaussian distributed random numbers (Andersen thermostat (AT))

Alternatively, new relative velocities can be taken from a Maxwell-Boltzmann distribution, i.e., a velocity $\boldsymbol{v}_i^{ran}$ is chosen for each particle with Gaussian distributed Cartesian components of zero mean and variance k_BT/m [46–48]. The new velocities are then given by

$$\boldsymbol{v}_i(t+h) = \boldsymbol{v}_{cm}(t) + \boldsymbol{v}_i^{ran} - \frac{1}{N_c} \sum_{i}^{N_c} \boldsymbol{v}_i^{ran}. \tag{51}$$

This algorithm conserves mass and momentum on the collision cell level. A canonical ensemble is simulated and no further thermalization is needed in non-equilibrium simulations, where there is viscose heating. From a numerical point of view, however, the calculation of the Gaussian random numbers is somewhat more time consuming compared to the rotations of the MPC-SRD version, hence the performance is slower [2].
In any case, the probability to find N_c particles in a cell is given by the Poisson distribution

$$P(N_c) = e^{-\langle N_c \rangle} \langle N_c \rangle^{N_c} / N_c! \,, \tag{52}$$

where $\langle N_c \rangle$ is the average number of the particles in a cell.
The described algorithms violate angular momentum conservation. In Refs. [21, 47, 48] algorithms are presented, which additionally preserve angular momentum.
More importantly, partition of space into collision cells implies a violation of Galilean invariance. This is most pronounced at low temperatures or small time steps, where the mean free path $h\sqrt{k_BT/m}$ of a particle is smaller than the cell size a. Then, the same particles repeatedly interact with each other in the same cell and thereby build up correlations. In a collision lattice moving with a constant velocity, other particles interact with each other, creating less correlations, which implies breakdown of Galilean invariance. In Refs. [49, 50], a random shift of the entire collisional grid is introduced to restore Galilean invariance. In practice, for sorting into collision cells, all particles are shifted by the same random vector with Cartesian components uniformly distributed in the interval $[-a/2, a/2]$. As a consequence, no reference frame is preferred.

Cell-Level Canonical Thermostat – As pointed out above, a thermostat is necessary for the MPC-SRD algorithm in any nonequilibrium situation, because the presence of external fields destroys energy conservation and a control mechanism has to be implemented to maintain temperature at the desired value (a brief review on existing thermostats is presented in Ref. [51]).

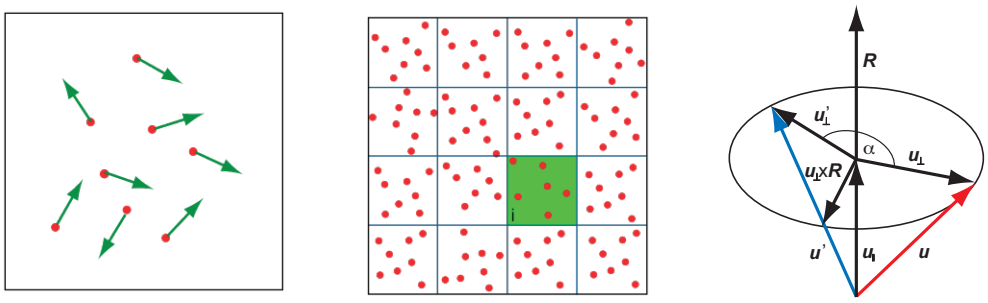


Fig. 3: *Illustration of the MPC-SRD dynamics: (left) streaming of particles; (middle) sorting of particles in collision lattice. (Right) Rotation of the vector* $\boldsymbol{u}$ *around the axis along* $\boldsymbol{R}$ *by the angle* α*.* $\boldsymbol{u}_\parallel$ *and* $\boldsymbol{u}_\perp$ *denote the vectors parallel and perpendicular to* $\boldsymbol{R}$*, respectively (cf. Appendix).*

A basic requirement of any thermostat is that it does not violate local momentum conservation, smear out local flow profiles, or distort the velocity distribution too much. A simple and efficient way to maintain a constant temperature is velocity scaling. For a homogeneous system, a single global scaling factor is sufficient. For an inhomogeneous system, e.g., fluids exposed to shear or Poiseuille flow, a local, profile-unbiased thermostat is required. Here, the relative velocities $\Delta\boldsymbol{v}_i = \boldsymbol{v}_i - \boldsymbol{v}_{cm}$ (49) can be scaled, before or after the rotation (velocity scaling exchanges with the rotation), i.e., $\Delta\boldsymbol{v}_i' = \kappa\Delta\boldsymbol{v}_i$, where κ is the scale factor.
In its simplest form, velocity scaling keeps the average kinetic energy $\langle E_k\rangle$ at the desired value $\langle E_k\rangle = 3(N_c-1)k_BT/2$. For a profile-unbiased *cell-level* scaling scheme, the scale factor is then given by

$$\kappa = \left(\frac{3(N_c-1)k_BT}{2E_k}\right)^{1/2}, \quad \text{where} \quad E_k = \frac{1}{2}\sum_{i=1}^{N_c} m\Delta\boldsymbol{v}_i^2 \tag{53}$$

is the kinetic energy of the particles within the particular cell. Note that the scale factor is different for every cell. This kind of temperature control corresponds to an isokinetic rather than isothermal, i.e., canonical ensemble [51].
A canonical ensemble is obtained for a suitable energy in the scale factor κ, which is different from the thermal average $\langle E_k\rangle = 3k_BT(N_c-1)/2$. Such a factor follows from the distribution function of the kinetic energy of a canonical ensemble [51]

$$P(E_k) = \frac{1}{E_k\Gamma(f/2)}\left(\frac{E_k}{k_BT}\right)^{f/2}\exp\left(-\frac{E_k}{k_BT}\right). \tag{54}$$

Here, $f = 3(N_c-1)$ denotes the degrees of freedom of the considered system and $\Gamma(x)$ is the gamma function. The distribution function $P(E_k)$ itself is denoted as gamma distribution. In the limit $f \to \infty$, the gamma distribution turns into a Gaussian function with the mean $\langle E_k\rangle = fk_BT/2$ and variance $f(k_BT)^2/2$.
To thermalize the velocities of the MPC fluid on the cell level, an energy E_k is taken from the distribution function (54) for every cell and time step and the velocities are scaled by the factor

$$\kappa = \left(\frac{2E_k}{\sum_{i=1}^{N_c} m\Delta\boldsymbol{v}_i^2}\right)^{1/2}. \tag{55}$$

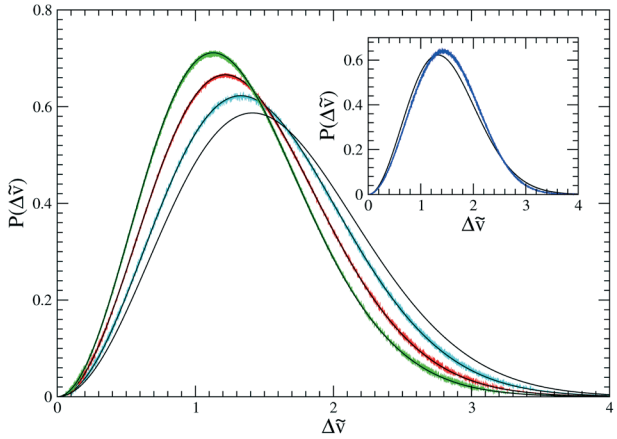


Fig. 4: *Distribution functions of the particle velocities $|\Delta\boldsymbol{v}|$ in a collision cell under shear flow for the average particle numbers $\langle N_c\rangle = 3$ (green), 5 (red), 10 (blue), and ∞ (black). The solid lines are determined using Eq. (57). $\Delta\tilde{v}$ is an abbreviation for $\Delta\tilde{v} = \Delta v/\sqrt{k_BT/m}$. The inset shows the distribution function for velocity scaling with the thermal energy $E_k = 3(N_c-1)k_BT/2$ for $\langle N_c\rangle = 10$ in comparison to the correct Maxwell-Boltzmann result (black) [51]. The simulation box size is $L = 30a$.*

For a fixed N_c, we then obtain the following distribution function for the relative velocity of a particle in a cell in the limit of a large number of MPC steps

$$P(\Delta\boldsymbol{v}, N_c) = \left(\frac{m}{2\pi k_BT(1-1/N_c)}\right)^{3/2} \exp\left(-\frac{m}{2k_BT(1-1/N_c)}\Delta\boldsymbol{v}^2\right). \tag{56}$$

However, the number of fluid particles in a cell is fluctuating in time. Thus, the actual distribution function is obtained by averaging Eq. (56) over the Poisson distribution (52)

$$P(\Delta\boldsymbol{v}) = \sum_{N_c=2}^{\infty} e^{-\langle N_c\rangle}\frac{\langle N_c\rangle^{N_c}}{N_c!}P(\Delta\boldsymbol{v}, N_c)/\left(1-(\langle N_c\rangle+1)e^{-\langle N_c\rangle}\right). \tag{57}$$

Figure 4 provides an example of velocity distributions of a MPC fluid under shear flow. Evidently excellent agrement is obtained between the simulation result and the theoretical expression.

4 Hydrodynamics and Multiparticle Collision Dynamics

In order to characterize the hydrodynamic properties of the MPC fluid we performed a series of simulations. We typically consider a three-dimensional system and apply periodic boundary conditions. We use the MPC parameters $\langle N_c\rangle = 10$, $\alpha = 130^o$, and $h = 0.1\sqrt{ma^2/k_BT}$. The box size L varies in the range $L/a = 20-150$, which corresponds to the total number of particles in the range $N \approx 10^5 - 3\times 10^7$. For an efficient simulation of the larger systems, we use a graphics processing units (GPU) accelerated MPC code [52].

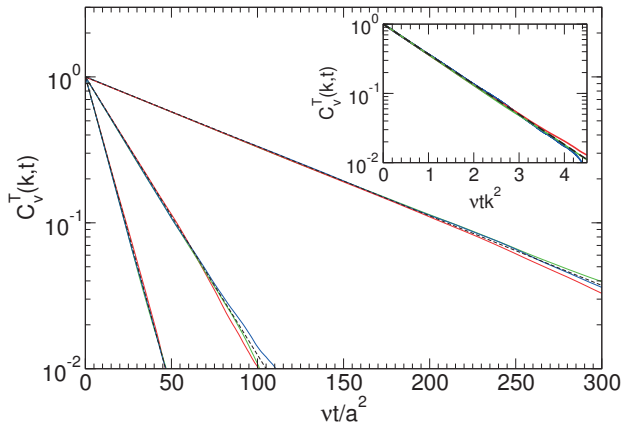


Fig. 5: *Transverse velocity autocorrelation functions of a MPC fluid for* $k = 2\pi n/L$*, with* $n = 1, 2, 3$ *(from right to left), and* $L/a = 60$*. The collision time steps are* $h/\sqrt{ma^2/(k_BT)} = 0.5$ *(blue),* 0.1 *(green), and* 0.01 *(red). The individual curves are hardly distinguishable. The corresponding theoretical prediction (25) is represented by a dashed line. Inset: universal dependence on* $\nu k^2 t$*.*

Velocity correlation function – To calculate velocity correlation functions in Fourier space, we perform the Fourier transformation

$$\boldsymbol{v}(\boldsymbol{k},t) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{v}_i(t) e^{i\boldsymbol{k}\cdot\boldsymbol{r}_i(t)} \tag{58}$$

of the MPC particle velocities, with the $\boldsymbol{k}$ vectors mentioned after Eq. (15). Figure 5 displays an example of the normalized transverse velocity autocorrelation function

$$C_v^T(t) = \left\langle \boldsymbol{v}^T(\boldsymbol{k},t)\cdot\boldsymbol{v}^T(-\boldsymbol{k},0)\right\rangle / \left\langle \boldsymbol{v}^T(\boldsymbol{k},0)\cdot\boldsymbol{v}^T(-\boldsymbol{k},0)\right\rangle . \tag{59}$$

The correlation functions decay exponentially, exactly as predicted by Eq. (25).

Long-time tail – Velocity correlation functions in real space are presented in Fig. 6, where

$$C_v(t) = \frac{m}{Nk_BT}\sum_{i=1}^{N}\left\langle \boldsymbol{v}_i(t)\cdot\boldsymbol{v}_i(0)\right\rangle . \tag{60}$$

The simulation data are well described by the theoretical expression Eq. (28), with Eqs. (25) and (26), over several decades in time. The theoretical approach even reproduces the oscillations due to sound at large times. However, we have to introduce an upper cut-off for the k values to describe the $C_v(t)$ at shorter times. The hydrodynamic description of the MPC fluid breaks down below a length scale somewhat larger than the collision cell size as indicated in Fig. 7 [24]. To achieve a good fit over a large time range, the maximum k value is set to $k_n = 2\pi n/60$ with $n = 16$ for $h/\sqrt{ma^2/(k_BT)} = 0.1$, which corresponds to the smallest length scale of $\approx 3.8a$. Studying large systems is quite time consuming. Hence, the long time behavior is illustrated in Fig. 6 (right) based on the analytical solution of Eq. (28). The dashed lines indicate negative

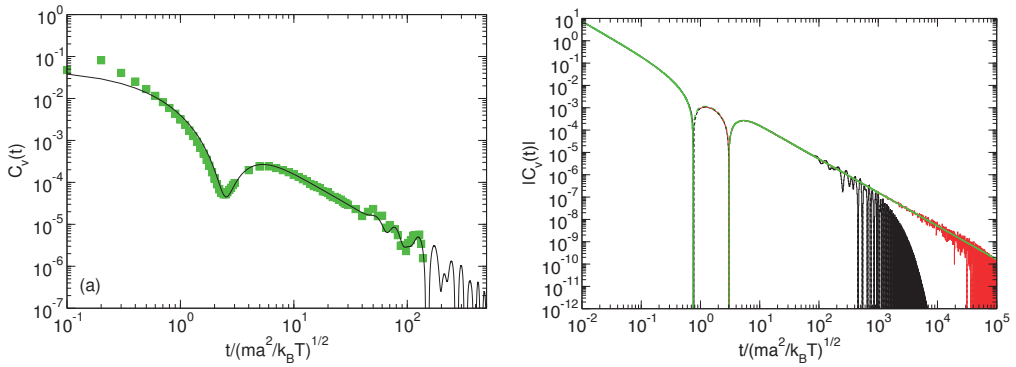


Fig. 6: *(Left) Magnitude of the velocity autocorrelation function (60) of a MPC fluid (symbols) for the systems size $L/a = 60$. The theoretical result (solid line) is obtained from Eq. (28). (Right) Theoretical velocity correlation functions [Eq. (28)] for the system sizes $L/a = 10^2$ (black), 10^3 (red), and ∞ (green). Negative parts of the correlation functions are indicated by dashed lines. For $t/\sqrt{ma^2/k_BT} \gtrsim 10^2$, the correlation function of the infinite system decays as $t^{-3/2}$ (long-time tail). In any case, the collision time step is $h/\sqrt{ma^2/(k_BT)} = 0.1$.*

correlation functions. They are a consequence of the oscillations of the longitudinal correlation function (26). This sound modes appear at $t/\sqrt{ma^2/k_BT} \approx 2$ for both, the simulations and the theoretical expression. This time roughly corresponds to the time by which sound has propagated over a collision cell, since the sound velocity for our isothermal MPC system is $c = \sqrt{k_BT/m}$. The correlation function at this time remains positive for the simulations (Fig. 6 (left)). The reason is that we took into account all possible k-values in Fig. 6 (right). As indicated above, the truncation of the k-sum modifies the correlation function at short times. With increasing system size, the sound oscillations shift to longer times – they appear at times $t > L/c$ as soon as sound has traversed the simulation box – and disappear for $L \to \infty$. For $t/\sqrt{ma^2/k_BT} \gtrsim 10^2$, the correlation function for the infinite system approaches asymptotically the long-time tail (29), where $C_v \sim t^{-3/2}$. The correlation functions of the finite-size systems deviate from the asymptotic behavior at some point in time. This is also a consequence of the finite system size, because there is a smallest k value $k_m = 2\pi/L$ corresponding to the largest available length scale. In the sum of k values in Eq. (28), k_m determines the long-time behavior and the correlation function finally decays exponentially for $t > L^2/(2\pi)^2\nu$.

Long-range hydrodynamic interactions – To illustrate the long-range character of hydrodynamic interactions, we consider the Oseen tensor $\mathbf{Q}^T(\boldsymbol{k})$ (32) in Fourier space, where the long-range nature of HI corresponds to a k^{-2} decay with increasing k values, or decreasing length scale. Figure 7 displays the amplitude $Q = 1/\eta k^2$ of the Oseen tensor for various MPC collision step sizes. For sufficiently small k values, $Q(\boldsymbol{k})$ shows the theoretically expected behavior. Above a collision-time step dependent value k, $Q(\boldsymbol{k})$ approaches a plateau. Applying the molecular chaos assumption, the asymptotic behavior can be calculated and is indicated by the horizontal lines [24]. Hence, below a certain length scale no hydrodynamic interactions are present anymore. This is a consequence of the lattice structure of the MPC method. An characteristic length scale λ_c, separating the hydrodynamic from the non-hydrodynamic regime, is

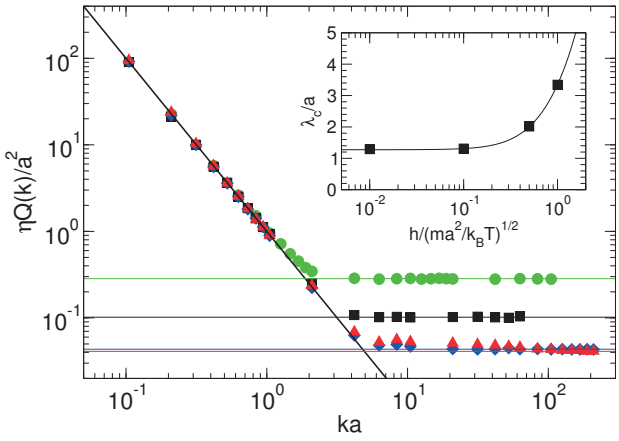


Fig. 7: *Dependence of* $Q(k) = 1/\eta k^2$ *on the wave number for the collision times* $h/\sqrt{ma^2/(k_BT)} = 0.01$ (▲), 0.1 (♦), 0.5 (■), *and* 1.0 (●). *The thick solid line indicates the dependence* $1/k^2$. *The horizontal lines are the theoretical predictions for the plateau values of* $\eta Q(k)$. *The inset shows the theoretical expectation for the characteristic length scale* λ_c *(solid line) and the values extracted from the simulations (squares).*

obtained by the intercept of the Oseen type dependence $Q(\boldsymbol{k}) = 1/\eta k^2$ with the asymptotic dependence $Q(\boldsymbol{k}) = Q_{\rm mc}(\boldsymbol{k}) = h/2\rho$, which yields $\lambda_c = \pi\sqrt{2\nu h}$. As indicated in Fig. 7, this prediction closely agrees with simulation results.
These results clearly show that MPC is a suitable simulation approach for fluids, since it captures their essential aspects. Naturally, a quantitative agreement with the solution of the Navier-Stokes equations can only be expected on sufficiently large length and times scales. In terms of length scale, this means scales larger than the collision cell size or, if the mean free path of a MPC particle is larger than a, larger than its mean-free path. As far as time is concerned, shear and sound waves should have propagated over one collision cell, i.e., $t > \tau_\nu \approx a^2/\nu$. This is consistent with the simulation results.

5 Multiparticle Collision Dynamics: Transport coefficients

Due to the simplicity of the algorithm, expressions for MPC fluid transport coefficients characterizing macroscopic laws can analytical be derived. In the following, the self-diffusion coefficient and the viscosity of the MPC fluid will be discussed. Other aspects are presented in Refs. [2, 45, 53].

5.1 Diffusion coefficient

The diffusion coefficient D of a particle i can be obtained from the Green-Kubo relation [2, 45, 50, 54]

$$D = \frac{h}{6}\left\langle \boldsymbol{v}_i(0)^2 \right\rangle + \frac{h}{3}\sum_{n=1}^{\infty}\left\langle \boldsymbol{v}_i(nh)\boldsymbol{v}_i(0)\right\rangle \tag{61}$$

for a discrete-time random system in three-dimensional space. $t_n = nh$ denotes the time of the nth collision. The average $\langle \ldots \rangle$ comprises both, averaging over the orientation of the rotation axis and the distribution of velocities. The two are independent. To evaluate the expression, the velocity autocorrelation function is required. An exact calculation of the MPC correlation function is difficult or even impossible, because it would imply that the full correlated dynamics of the particles can analytically be resolved. However, an approximate expression can be derive by applying the molecular chaos assumption, i.e., by assuming that the velocities of different particles are uncorrelated $\langle \boldsymbol{v}_i(t) \cdot \boldsymbol{v}_j(t') \rangle = 0$, $\forall\ i \neq j$. Then, the velocity autocorrelation function reduces to

$$\langle \boldsymbol{v}_i(nh) \cdot \boldsymbol{v}_i(0) \rangle = (1-\gamma)^n \left\langle \boldsymbol{v}(0)^2 \right\rangle, \quad \text{with} \quad \gamma = \frac{2(1-\cos\alpha)}{3\langle N_c \rangle} \left(e^{-\langle N_c \rangle} + \langle N_c \rangle - 1 \right). \quad (62)$$

Hence, the correlation function decays exponentially, which is certainly not correct as shown in Fig. 6. The diffusion coefficient itself follows as

$$D = \frac{h \langle \boldsymbol{v}_i(0)^2 \rangle}{3} \left(\frac{1}{\gamma} - \frac{1}{2} \right) = \frac{h k_B T}{m} \left(\frac{1}{\gamma} - \frac{1}{2} \right). \quad (63)$$

The comparison with simulations shows that D underestimates the diffusion coefficient for $h/\sqrt{ma^2/k_BT} < 1$. For $h/\sqrt{ma^2/k_BT} = 0.1$, $\langle N_c \rangle = 10$, and $\alpha = 130^\circ$, the simulated value is about 30% larger. For $h/\sqrt{ma^2/k_BT} \gg 1$, the molecular chaos assumption applies and the analytical expression agrees well with simulation results.

5.2 Viscosity

The shear viscosity is one of the most important properties of complex fluids. In particular, it characterizes their non-equilibrium behavior, e.g., in rheology. Various ways have been suggested to obtain an analytical expression for the viscosity of a MPC fluid. In Refs. [2,50,54–56], linear hydrodynamic equations (Navier-Stokes equation) and Green-Kubo relations are exploited. Alternatively, non-equilibrium simulations can be performed and transport coefficients are obtained from the linear response to an imposed gradient [57]. The two approaches are related by the fluctuation-dissipation theorem.

Stress tensor – In simple shear flow, with the velocity field $v_x = \dot{\gamma} y$, where v_x is the fluid flow field along the x-direction (flow direction), y the gradient direction, and $\dot{\gamma}$ the shear rate, the viscosity η is related to the macroscopic stress tensor via

$$\sigma_{xy} = \eta \dot{\gamma}. \quad (64)$$

Hence, an expression is required for the stress tensor to either derive η analytically and/or to determine it in simulations. In Refs. [58, 59], the kinetic theory moment method has been applied to derive an analytical expression.
Here, the stress tensor derived by the virial theorem is presented [57]. This derivation starts with the equations of motion of a particle and leads to the instantaneous stress tensor

$$\sigma^i_{\alpha\beta} = -\frac{1}{V} \sum_{i=1}^{N} m \hat{v}_{i\alpha} \hat{v}_{i\beta} - \frac{1}{Vh} \sum_{i=1}^{N} \Delta p_{i\alpha} r_{i\beta}\,, \quad (65)$$

except of the component σ_{xy}, which reads as

$$\sigma_{xy}^{i} = -\frac{1}{V}\sum_{i=1}^{N} m\hat{v}_{ix}\hat{v}_{iy} - \frac{\dot{\gamma}h}{2V}\sum_{i=1}^{N} mv_{iy}^{2} - \frac{1}{Vh}\sum_{i=1}^{N}\Delta p_{ix}r_{iy}. \tag{66}$$

The hat ($\hat{v}_{i\alpha}$) indicates velocities after streaming and all velocities and positions are inside the primary box of the underlaying periodic simulation box, i.e., Lees-Edwards periodic boundary conditions are applied [44, 57] (see Sec. 7.1). The superscript i indicates the instantaneous internal stress tensor. In addition, an alternative expression, denoted as external stress, can be defined [57].

Analytical expression for viscosity – The above expressions for the stress tensor are independent of any particular collision rule. The viscosity of a system, however, depends on the applied collision procedure. Analytical expressions for the viscosity of a MPC-SRD fluid have been derived by various approaches [2, 14, 45, 48, 49, 56–59].
In simple shear flow, the viscosity η is determined by Eq. (64), where the (macroscopic) stress tensor follows from $\sigma_{xy} = \langle\sigma_{xy}^{i}\rangle_{s}$ – here, S indicates averages over MPC steps. For a MPC fluid, the stress tensor is composed of a kinetic and collisional contribution, i.e, $\sigma_{xy} = \sigma_{xy}^{\mathrm{kin}} + \sigma_{xy}^{\mathrm{col}}$, which implies that the viscosity $\eta = \eta_{\mathrm{kin}} + \eta_{\mathrm{col}}$ consists of a kinetic η_{kin} and collisional η_{col} part too. The two contributions are conveniently obtained from the stress tensor (66). The detailed calculations presented in Refs. [57, 60] yield

$$\eta_{\mathrm{kin}} = \frac{Nk_BTh}{V}\left[\frac{5\langle N_c\rangle}{(\langle N_c\rangle - 1)(4 - 2\cos\alpha - 2\cos(2\alpha))} - \frac{1}{2}\right], \tag{67}$$

$$\eta_{\mathrm{col}} = \frac{Nma^2}{18Vh}(1-\cos\alpha)\left(1 - \frac{1}{\langle N_c\rangle}\right). \tag{68}$$

It has to be emphasized that the molecular chaos assumption is used in the derivation of the expression for η_{kin}. Hence, a certain deviation between the theoretical expressions and simulation results is expected. However, recent simulations [61] indicate that the theoretical expression $\eta = \eta_{\mathrm{kin}} + \eta_{\mathrm{col}}$ describes simulation results within an accuracy better than 2%.
Figure 8 displays results for viscosities determined for a MPC-SRD fluid. The good agreement between theory and simulations is evident. The viscosity increases with increasing and decreasing h. The two limits correspond to different system behaviors. For large h, the viscosity is dominated by the kinetic part. This corresponds to the limit of a gas-like system. At small h, the collisional contribution dominates, which corresponds to the limit of fluid-like behavior.

Schmidt number – A convenient measure of the importance of hydrodynamics is the Schmidt number $Sc = \nu/D$, where $\nu = \eta/(m\langle N_c\rangle)$ is the kinematic viscosity [62]. Evidently, Sc is the ratio between momentum transport and mass transport. As is known, this number is smaller than but on the order of unity for gases, while in fluids, like water, it is on the order of 10^2 to 10^3. A prediction for the Schmidt number of a MPC fluid can be obtained from the analytical expressions (67) and (68) for the viscosity, and the diffusion coefficient (63). In Fig. 8, the theoretical prediction for Sc is displayed for different values of the rotation angle. This shows that Sc becomes considerably larger than unity for $h \to 0$. In fact, Sc increases like $1/h^2$ as soon as the collisional viscosity dominates over the kinetic viscosity. Hence, we obtain fluid-like behavior in the limit of small collision-time steps.

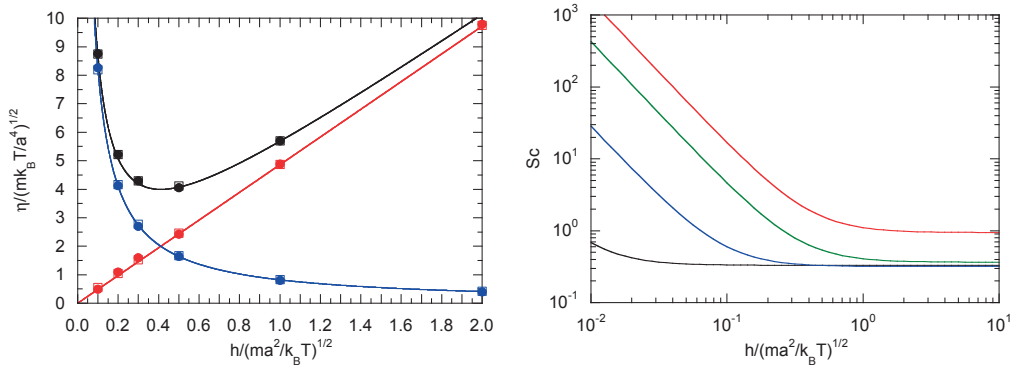


Fig. 8: *(Left) Simulation results (symbols) for the total viscosity (black) with its kinetic (red) and collisional (blue) contributions determined via the internal (bullets) and external (open squares) stress tensors [57]. The analytical results [Eqs. (67) and (68)] are presented by solid lines. (Right) Theoretical Schmidt numbers as function of the collision time step h for the rotation angles $\alpha = 15^\circ$ (black), 45° (blue), 90° (green), and 130° (red). The mean particle number is $\langle N_c\rangle = 10$.*

6 MPC: Embedded object and boundary conditions

A very simple procedure for coupling embedded objects such as colloids or polymers to a MPC fluid has been proposed in Refs. [63–66]. In this approach, a colloidal particle or a polymer is composed of point particles which are connected by a bond potential to maintain the respective shape. To couple the object to the MPC fluid, the individual particles (monomers) participate in the MPC collision. If particle (monomer) k has mass M and velocity $\boldsymbol{V}_k$ the center-of-mass velocity of all particles (MPC and monomers) in a collision cell is

$$\boldsymbol{v}_{cm} = \frac{\sum_{i=1}^{N_c} m\boldsymbol{v}_i + \sum_{k=1}^{N_m^c} M\boldsymbol{V}_k}{mN_c + MN_m^c}, \tag{69}$$

where N_m^c is the number of monomers in the collision cell. A stochastic rotation of the relative velocities of both the fluid particles and embedded monomers is then performed in the collision step, which leads to an exchange of momentum between them. The dynamics of the monomers is typically treated by molecular dynamics simulations (MD), applying the velocity Verlet integration scheme [44, 67]. Hence, the new monomer momenta are used as initial conditions for the the subsequent streaming step (MD) of duration h involving several MD steps, since the MD simulation time step is typically smaller than h. In this approach, the average mass of fluid particles per cell $m\langle N_c\rangle$ should be of the order of the total monomer mass MN_m^c. This corresponds to a neutrally buoyant object which responds quickly to the fluid flow but is not kicked around too violently. It is also important to note that the average number of monomers per cell $\langle N_m\rangle$ should be on the order of unity to properly resolve HI between them. On the other hand, the average bond length in a semiflexible polymer or rodlike colloid should also not be much larger than the cell size a, in order to capture the anisotropic friction of rodlike molecules due to HI (which leads to a twice as large perpendicular than parallel friction coefficient for long

stiff rods [20, 22]), and to avoid an unnecessarily large ratio of the number of fluid-to-solute particles. Hence, the average bond length should be of order a.

To accurately resolve the local flow field around a hard-sphere colloid, methods have been proposed which exclude fluid-particles from the interior of the colloid and mimic slip [14, 68] or no-slip boundary conditions [2, 69–73]. No-slip boundary conditions are modeled by the bounce-back rule. Here, the relative velocity of a fluid particle $\tilde{\boldsymbol{v}}_i = \boldsymbol{v}_i - \boldsymbol{v}_s$, where $\boldsymbol{v}_s$ is the velocity of the surface point $\boldsymbol{r}_s$ hit by the particle i, is inverted from $\tilde{\boldsymbol{v}}_i$ to $-\tilde{\boldsymbol{v}}_i$, when it intersects the surface of an impenetrable particle, e.g., a colloid, a blood cell, or a wall. The velocity $\boldsymbol{v}_s$ may comprise translational and rotational components, i.e, $\boldsymbol{v}_s(\boldsymbol{r}_s) = \boldsymbol{V} + \boldsymbol{\omega} \times (\boldsymbol{r}_s - \boldsymbol{R})$, where $\boldsymbol{V}$ and $\boldsymbol{R}$ are the center-of-mass velocity and position of the colloid, respectively, and $\boldsymbol{\omega}$ is its rotational frequency.

Since walls or surfaces will generally not coincide with the collision cell boundaries, in particular due to a random shift, the simple bounce-back rule fails to guarantee no-slip boundary conditions. To establish no-slip boundary conditions or at least reduce slip as far as possible, the following procedure has been suggested [69]: For all collision cells that are intersected by walls, fill the wall part of the cell with a sufficient number of virtual (phantom) particles in order to ensure that the total number of particles is equal to $\langle N_c \rangle$. The velocities of the wall particles are taken from a Maxwell-Boltzmann distribution with an appropriate mean and variance $k_B T/m$. For illustration, I will focus on a planar wall at rest in the following. The more general case of moving colloids is addressed in Refs. [71–73].

Since the sum of Gaussian random numbers is also Gaussian distributed, the velocities of the individual virtual particles need not be determined explicitly in simple geometries, it suffices to determine a momentum $\boldsymbol{p}$ from a Maxwell-Boltzmann distribution with zero mean and variance $mN_p k_B T$, where $N_p = \langle N_c \rangle - N_{sc}$ is the number of virtual particles corresponding to the partially filled cell of N_{sc} particles. The center-of-mass velocity of the cell is then

$$\boldsymbol{v}_{cm} = \frac{1}{m \langle N_c \rangle} \left(\sum_{i=1}^{N_{sc}} m\boldsymbol{v}_i + \boldsymbol{p} \right). \tag{70}$$

Results for a Poiseuille flow obtained by this procedure are in good agreement with the correct parabolic flow profile [51, 69].

However, this does not completely prevent slip, because the average center-of-mass position of all particles in a collision cell – including the virtual particle – does not coincide with the wall. In order to further reduce slip, the following modification of the original approach has been proposed [57]. To treat a surface cell on the same basis as a cell in the bulk, i.e., the number of particles satisfies the Poisson distribution with the average $\langle N_c \rangle$, we take fluctuations in the particle number into account by adding N_p virtual particles to every cell intersected by a wall such that $\langle N_p + N_{sc} \rangle = \langle N_c \rangle$. There are various ways to determine the number N_p. For a system with parallel walls, we suggest to use the number of fluid particles in the opposite surface cell, i.e., the opposing surface cell cut by the other wall. The average of the two numbers is equal to $\langle N_c \rangle$. Alternatively, N_p can be taken from a Poisson distribution with average $\langle N_c \rangle$ accounting for the fact that there are already N_c particles in the cell. Now, the center-of-mass velocity of the particles in a boundary cell is

$$\boldsymbol{v}_{cm} = \frac{1}{m(N_{sc} + N_p)} \left(\sum_{i=1}^{N_{sc}} m\boldsymbol{v}_i + \boldsymbol{p} \right). \tag{71}$$

The momentum $\boldsymbol{p}$ of the effective virtual particle is obtained as described above.

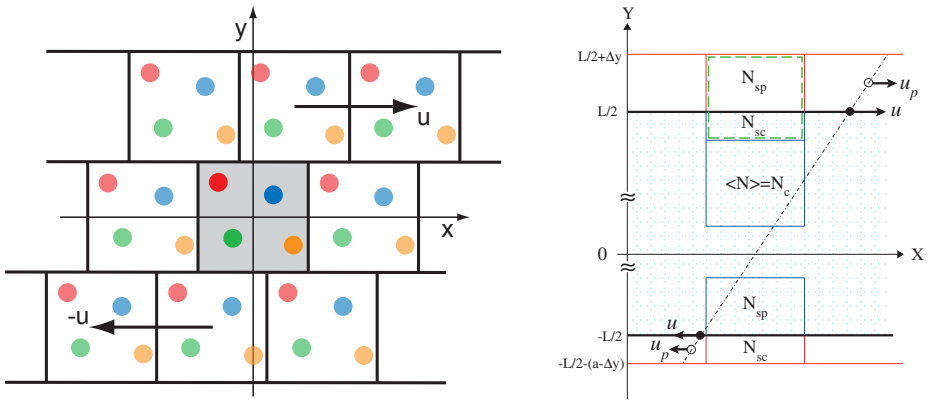


Fig. 9: *(Left) Lees-Edwards homogeneous shear boundary conditions. The primary box is highlighted in gray. The opaque particles are periodic images of the particles of the primary box. The upper layer is moving with the velocity $u = \dot{\gamma} L_y$ to the right, and the bottom layer to the left. Note that the shear velocity is zero in the center of the primary box [22]. (Right) Illustration of the random shift and the distribution of particles in collision cells intersected by the walls. The walls are located at $y = \pm L/2$. Under shear, the walls move with the velocities $u = \dot{\gamma} L/2$. A phantom particle is located in the center of every truncated part of a collision cell at $(L + \Delta y)/2$ and $-(L + a - \Delta y)/2$, respectively. They move with the velocities $u_p = \dot{\gamma}(L+\Delta y)/2$ and $u_p = -\dot{\gamma}(L+a-\Delta y)/2$, respectively. The dashed-dotted line indicates the linear velocity profile.*

7 MPC: External fields

7.1 Shear flow

To impose shear flow on a periodic MPC fluid system, Lees-Edwards boundary conditions are applied [44,74]. As illustrated in Fig. 9 (left), the infinite periodic system is subject to a uniform shear in the xy-plane [22]. The layer of simualtion boxes with the primary box (shaded) is stationary, whereas the layer above moves with the velocity $u = \dot{\gamma} L$ to the right and the layer below with $-u$ to the left. The corresponding further layers move with the respective integral multiple of u. However, these further layers are not required in practice. Whenever a MPC particle leaves the primary box, it is replaced by its periodic image. This avoids build-up of a substantial difference in the x-coordinates [22].

When walls are present, shear flow can be imposed by the opposite displacement of the confining walls with the velocities $u = \pm\dot{\gamma} L/2$ as indicated in Fig. 9 (right) (the reference frame is fixed in the center of the simulation box). Here, shear is imposed in two ways, by applying bounce-back boundary conditions, i.e., the momentum of a particle changes as $\Delta \boldsymbol{p}_i = -2m\boldsymbol{v}_i + 2m\boldsymbol{u}$ ($\boldsymbol{u} = (u, 0, 0)^T$) during streaming, and by collisions with virtual wall particles [57]. The momentum of a virtual particle is determined from a Boltzmann distribution, with the mean value along the flow direction

$$p_u = mN_p \left(u + \frac{\dot{\gamma}}{2} \Delta y \right) \tag{72}$$

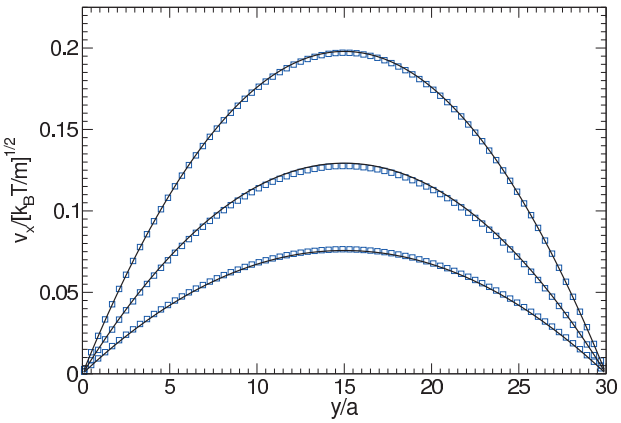


Fig. 10: *Poiseuille-flow velocity profiles for the collision-time steps* $h/\sqrt{ma^2/k_BT} = 1.0, 0.1,$ *and* 3.0 *(symbols, top to bottom). The solid lines are fits of Eq. (73).*

for a surface at $+L/2$. For the surface at $-L/2$, $u \to -u$ and $\Delta y \to -(a - \Delta y)$ for a given random shift. Here, Δy is the fraction of the wall-truncated collision cell insight the wall and N_p denotes the number of virtual particles, where $N_p = N_{sp}$ for the upper wall and $N_p = N_{sc}$ for the lower wall [57].

7.2 Poiseuille flow

A parabolic flow profile of a fluid confined between walls is obtained by a constant pressure gradient or a uniform body force, e.g., gravitational force, combined with non-slip boundary conditions. For two planer walls parallel to the xz-plane at $y = 0$ and $y = L$, the Stokes equation yields the velocity profile

$$v_x(y) = \frac{4v_{\max}y(L-y)}{L^2}, \quad \text{with} \quad v_{\max} = \frac{m\langle N_c\rangle gL^2}{8\eta a^3} = \frac{gL^2}{8\nu}. \tag{73}$$

Here, $\varrho g = m\langle N_c\rangle g/a^3$ is the gravitational (volume) force density [69].

In MPC simulations, a parabolic flow profile is obtained in a similar manner by confining the fluid between walls and applying periodic boundary conditions parallel to them. Every fluid particle is exposed to the gravitational force $F_x = mg$ along the x-direction. Naturally, other channel geometries, such as square channels [75] or capillaries [76–78] can be considered. Then, the particle velocities and positions are updated according to

$$\begin{aligned} \hat{v}_{i\beta}(t+h) &= v_{i\beta}(t) + gh\delta_{\beta x}, \\ \hat{r}_{i\beta}(t+h) &= r_{i\beta}(t) + v_{i\beta}(t)h + \frac{1}{2}gh^2\delta_{\beta x} \end{aligned} \tag{74}$$

in the streaming step. The bounce-back rule has to be adjusted too. This is simply done after the streaming step (74) is complete. The velocities and positions of the particles which penetrated

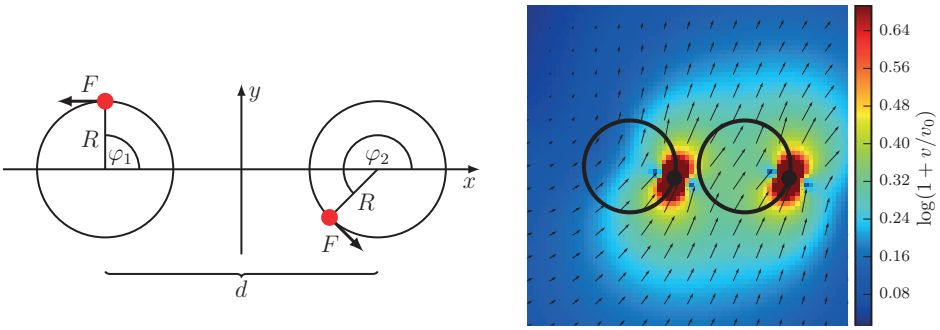


Fig. 11: *(Left) Model system of hydrodynamically coupled rotors. The beads move along fixed circular trajectories in the xy-plane, each driven by a constant tangential force F. (Right) Flow field in the stationary synchronized state for the Péclet number $Pe = 120$. The reddish colors indicate the dipolar character of the flow field. The flow field is calculated from the hydrodynamic approach without thermal fluctuations [21].*

a wall are corrected according to

$$\begin{aligned}\hat{v}_{i\beta}(t+h) &= -v_{i\beta}(t+h) + 2g\Delta h_i \delta_{\beta x},\\ \hat{r}_{i\beta}(t+h) &= r_{i\beta}(t+h) - 2v_{i\beta}(t+h)\Delta h_i \delta_{\beta x} + 2g\Delta h_i^2 \delta_{\beta x}.\end{aligned} \tag{75}$$

The time Δh_i, during which the particle moves insight the wall, follows from the dynamics along the y-direction: $\Delta h_i = [\hat{r}_{iy}(t+h) - L\Theta(\hat{r}_{iy}(t+h) - L)]/v_{iy}(t+h)$.
Figure 10 shows examples of flow profiles for various collision step sizes. Evidently the parabolic shape is reproduce well. By the fit of Eq. (73), the viscosity of the fluid is obtained.

8 Applications

As examples for the MPC approach, two aspects will be briefly touched, namely (i) the synchronization of the rotational motion of two microrotors by time-dependent hydrodynamic interactions and (ii) the hydrodynamic properties of discrete-particle models for colloids.

8.1 Hydrodynamic synchronization of microrotors

Synchronization of motion is a common phenomenon in nonlinear many-particle systems, and thus appears in a broad range of physical, biological, engineering, and social systems [79, 80]. The phenomenon appears at all length scales from atoms to macroscopic bodies. For microswimmers, synchronization is fundamental for the coordinated cyclic motion of cilia and flagella (see also lecture F.7 "Microswimmers"). The synchronous beating of the two flagella of *Chlamydomonas* causes straight swimming, while asynchronous beating implies tumbling motion [81–86]. The helical flagella of bacteria, like *E. coli*, synchronize their rotational motion during bundling [87–91]. Multiciliated and multiflagellated microorganisms such as unicellular Paramecia [92] or Volovoxes [93] exhibit metachronal waves (MCW) [94]. Here, synchronization is essential for microswimmer motility. Furthermore, coordinated flagellar motion plays a major role in eukaryotes [95, 96], where they transport fluid in the respiratory system in form of

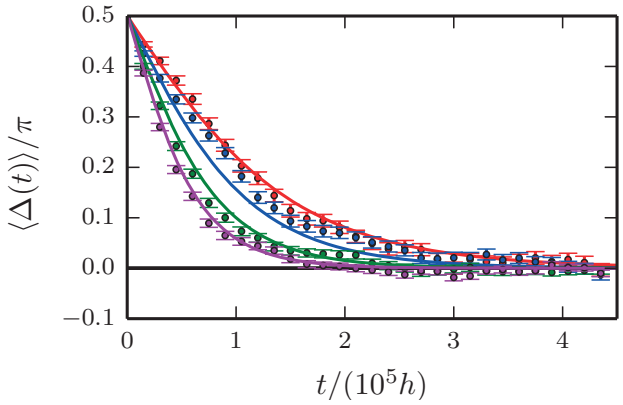


Fig. 12: *Average phase-angle difference $\Delta(t)$ for the Péclet numbers $Pe = 120,\ 140,\ 160$, and 180 (top to bottom). Symbols represent simulation results and the solid lines are obtained by numerical integration of Eq. (79) including transversal and longitudinal modes of the hydrodynamic tensor [21].*

cilia [97], are involved in cellular communications [98], and even determine the morphological left-right asymmetry in the embryo [99].

From a microswimmer point of view, a rotor is the most simple description of a beating cilium [100]. For example, its endpoint moves along a closed trajectory, which might be approximated by a circle. The rotor model is illustrated in Fig. 11. The two beads of radius R_H move along circles of radius R, each driven by an active force $\boldsymbol{F}_i$. The circles are centered at $\boldsymbol{R}_i^0 = (-1)^i (d/2)\boldsymbol{e}_x$ $(i = 1, 2)$, where $\hat{\boldsymbol{e}}_x$ is the unit vector along the x-axis and d the center-to-center distance. The trajectories of the bead centers can be expressed as

$$\boldsymbol{R}_i(t) = \boldsymbol{R}_i^0 + R(\cos\varphi_i(t), \sin\varphi_i(t), 0)^T, \tag{76}$$

in terms of the phase angles $\varphi_i(t)$. The driving forces $\boldsymbol{F}_i(t) = F\hat{\boldsymbol{t}}_i(t)$ are of equal magnitude and point along the tangents $\hat{\boldsymbol{t}}_i(t) = (-\sin\varphi_i(t), \cos\varphi_i(t), 0)^T$.
The coupling of the beads of mass M with the MPC fluid is established in the collision step as described by Eq. (69). The velocity $\boldsymbol{V}_i(t+h)$ of a bead after a collision follows from Eq. (49). Hence, the phase angles evolve as

$$\varphi_i(t+h) = \varphi_i(t) + \dot{\varphi}_i(t)h + \frac{1}{2}\left(\frac{F}{MR}\right)h^2 \tag{77}$$

between MPC collisions. In a collision, the angular velocities change. The new values after a collision follow from the relation

$$\dot{\varphi}_i(t+h) = \frac{1}{R}\hat{\boldsymbol{t}}_i(t+h)\cdot\boldsymbol{V}_i(t+h). \tag{78}$$

Figure 12 shows average phase-angle differences $\Delta(t) = \varphi_1(t) - \varphi_2(t)$ of co-rotating beads. The averages $\langle\Delta(t)\rangle$ indeed show synchronization of the bead rotation despite strong thermal fluctuations for the various Péclet numbers Pe, where Pe is defined as $Pe = FR/k_BT$ [21].

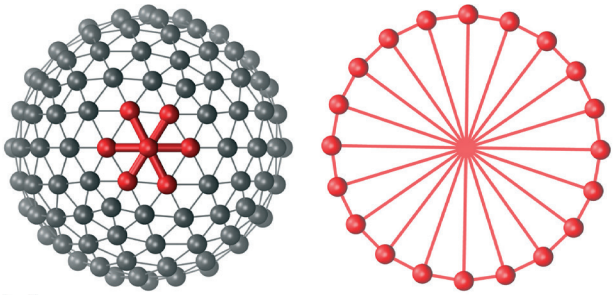


Fig. 13: *Spherical colloid of radius $R = 3a$ composed of discrete mass points with the surface density $\sigma a^2 = 1.43$. For illustration, the bonds of a particle and its nearest neighbors are highlighted in red. Bonds in an equatorial plane are connecting diametrically with opposite particles (right) [66].*

The following approximate analytical expressions for the dynamics of the phase angles φ_i can be derived in the absence of thermal fluctuations and within the mean-field approximation $\boldsymbol{R}_2(t) - \boldsymbol{R}_1(t') \approx \boldsymbol{d} = d\hat{\boldsymbol{e}}_x$ [21,33]:

$$\dot{\varphi}_i = \omega + \frac{F}{R} \sum_{j \neq i} \int_0^t dt' \, \hat{\boldsymbol{t}}_i(t) \cdot \mathbf{Q}(d\boldsymbol{e}_x, t - t') \, \hat{\boldsymbol{t}}_j(t'), \tag{79}$$

where $\omega = F/\gamma R$ is the intrinsic angular frequency and $\gamma = 6\pi\eta R_H$ Stokes' friction coefficient of a bead.
As is evident from the Fig. 12, the simulation results compare well with the theoretical predictions of Eqs. (79) when taking into account compressibility effects, although there are small differences for $Pe = 140$ and 160.
The flow field

$$\boldsymbol{v}(\boldsymbol{r}, t) = F \sum_{i=1}^{2} \int_0^t \mathbf{Q}(\boldsymbol{r} - \boldsymbol{r}_i(t'), t - t') \hat{\boldsymbol{t}}_i(t') \, dt' \tag{80}$$

is illustrated in Fig. 11 (right). The velocity profile decays very fast with separation from a beat. Thereby, it is anisotropic – of Stokeslet shape – with a slower decay in the tangential forward and backward direction. The synchronous state is preferable, because it minimizes dissipation.

8.2 Hydrodynamics of discrete-particle models

As discussed in Sec. 6, coupling of the MPC fluid with objects composed of point particles is particular efficient. Here, the question arises to which extend solid and impenetrable objects, such as hard-sphere colloids, can be modelled by a point-particle representation. The advantage of a point-particle model lies not only in its computational efficiency, but also rather complex shapes can easily be constructed and be coupled with the MPC fluid [66].

The colloid model is displayed in Fig. 13. The set of N_s point particles of mass M are distributed over a spherical shell of radius R and are tightly linked with each other to maintain a nearly rigid

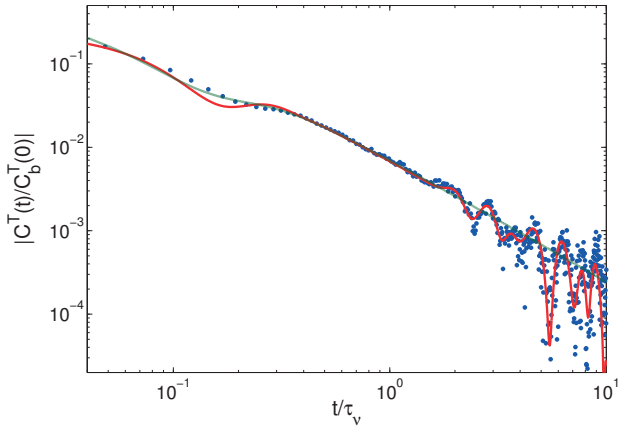


Fig. 14: *Colloid center-of-mass velocity autocorrelation function (blue dots) compared with the analytical results for a neutrally buoyant hard-sphere colloid (green) and the analytical result Eq. (81) (red) for* $h/\sqrt{ma^2/k_BT} = 0.1$ *and* $K = 5 \times 10^3 k_BT/a^2$. *Time is scaled by* $\tau_\nu = 4R^2/\nu$. $C^T(t)$ *is normalized by the value* $C_b^T(0) = 3k_BT/M_c$ *of a neutrally buoyant hard-sphere colloid. For details, see Ref. [66].*

shape. The actual number of points depends on the radius of the colloid. A suitable number of points has to be chosen in order to achieve proper hydrodynamic behavior. We start out with a truncated icosahedron with a particle at every vertex. Higher surface densities are constructed by dividing the faces of the icosahedron into equilateral triangles [66]. The mass points added at the polyhedrons are displaced radially from the center until they lie on a shell of radius R. The mass points are connected with their nearest neighbors and with the diametrically opposite point particle via the harmonic potential $U = K(r - r_0)^2/2$, where $r = |\boldsymbol{r}|$ is the distance between two particles and $\boldsymbol{r}$ the respective difference vector. The equilibrium distance r_0 depends on the particular pair and is determined by the above described construction principle. The dynamics of a colloid particle is described by Newtons equations of motion, as discussed in Sec. 6.

The colloid transverse center-of-mass velocity autocorrelation function is displayed in Fig. 14. We find very good agreement between the theoretical correlation function of a buoyant hard-sphere colloid, i.e., its density is equal to the density of the fluid, and simulation results, as well as with the expression

$$C^T(t) = \frac{1}{N_s}\frac{k_BT}{V}\sum_{\boldsymbol{k}} S(\boldsymbol{k})\left(2Q^T(\boldsymbol{k},t) + Q^L(\boldsymbol{k},t)\right), \tag{81}$$

with the structure factor

$$S(\boldsymbol{k}) = \frac{1}{N_s}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s}\langle e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_i - \boldsymbol{R}_j)}\rangle, \tag{82}$$

based on a point-particle model [32, 66]. Equation (81) is rather similar to Eq. (28), except of the structure factor, which accounts for the geometry.

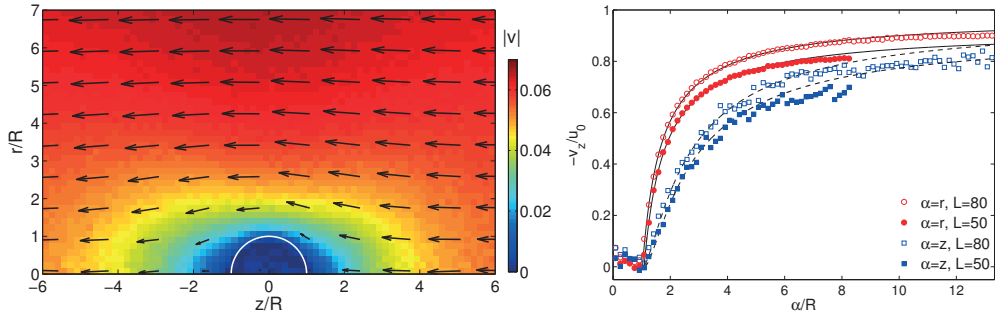


Fig. 15: *Fluid velocity field in the rest frame of a sedimenting colloid. Arrows indicate the direction of the velocity field $\boldsymbol{v}/|\boldsymbol{v}|$. The magnitude of the field $|\boldsymbol{v}|$ is color-coded. Cylindrical coordinates are used, with the cylinder axis along the z-axis of the Cartesian reference frame (sedimentation direction) and the radial coordinate r. Comparison of the fluid velocity component v_z obtained from simulation (symbols) and calculated according to Eq. (83) (lines) along the sedimentation direction z (blue, squares) and along the radial direction r (red, circles/bullets) for the system sizes $L = 80a$ (open symbols) and $L = 50a$ (closed symbols). System parameters are $\sigma a^2 = 1.43$ and $h = 0.1t_0$. The velocities are scaled by the value $u_0 = f/(6\pi\eta R)$.*

The flow field generated by a sedimenting colloid is shown in Fig. 15 (left). The strength of the velocity field along the sedimentation and radial directions are displayed in Fig. 15 (right). The numerical results are compared with the analytical expression

$$\boldsymbol{v} = \boldsymbol{u} - \frac{3}{4}R_H\frac{\boldsymbol{u} + \hat{\boldsymbol{r}}(\boldsymbol{u}\cdot\hat{\boldsymbol{r}})}{|\boldsymbol{r}|} - \frac{1}{4}R_H^3\frac{\boldsymbol{u} - 3\hat{\boldsymbol{r}}(\boldsymbol{u}\cdot\hat{\boldsymbol{r}})}{|\boldsymbol{r}^3|}, \tag{83}$$

for the flow field around a sphere with no-slip boundary conditions in the Stokes regime [16]. Here, $\boldsymbol{u}$ is the fluid velocity at infinity, $\boldsymbol{r}$ is the position vector from the center of the colloid, and $\hat{\boldsymbol{r}} = \boldsymbol{r}/|\boldsymbol{r}|$. We assume that the flow field is $u = f/(6\pi\eta R_H)$ at infinity. The simulation data agree well with the theoretical expression. In particular, the flow velocity in the colloid interior is very close to zero and the no-slip boundary condition is very well satisfied.

In summary, the discrete particle model provides a valuable alternative to a solid colloid as far as hydrodynamic aspects are concerned, with a considerably reduced numerical effort. This is particularly interesting for more complex structures such as coarse-grained protein models with mobile domains.

9 Conclusions

Mesoscale hydrodynamic simulation approaches are extremely valuable tools to study soft matter systems. Specifically, the MPC approach developed into a versatile tool to study hydrodynamic properties of complex fluids since it has been introduced in 1999 [13]. By now, several collision algorithms have been proposed and employed, and the method has been generalized to describe multi-phase flows and viscoelastic fluids. A major advantage of the algorithm is

its easy coupling to the dynamics of embedded particles using a hybrid MPC-MD simulations approach. Results of such studies are in excellent quantitative agreement with both theoretical predictions and results obtained using other simulation techniques. In the future, we will see more applications of the method in non-equilibrium and driven soft-matter systems; specifically for systems where thermal fluctuations play a major role. Here, the full advantage of the method can be exploited, because the interactions of colloids, polymers, and membranes with the mesoscale fluid can be treated on the same basis.

Appendices

Collision: Velocity rotation

The rotations can be realized in different ways. On the one hand, the rotation matrix

$$\mathbf{R}(\alpha) = \begin{pmatrix} \mathcal{R}_x^2 + (1-\mathcal{R}_x^2)c & \mathcal{R}_x\mathcal{R}_y(1-c) - \mathcal{R}_z s & \mathcal{R}_x\mathcal{R}_z(1-c) + \mathcal{R}_y s \\ \mathcal{R}_x\mathcal{R}_y(1-c) + \mathcal{R}_z s & \mathcal{R}_y^2 + (1-\mathcal{R}_y^2)c & \mathcal{R}_y\mathcal{R}_z(1-c) - \mathcal{R}_x s \\ \mathcal{R}_x\mathcal{R}_z(1-c) - \mathcal{R}_y s & \mathcal{R}_y\mathcal{R}_z(1-c) + \mathcal{R}_x s & \mathcal{R}_z^2 + (1-\mathcal{R}_z^2)c \end{pmatrix} \tag{84}$$

can be used, with the unit vector $\boldsymbol{\mathcal{R}} = (\mathcal{R}_x, \mathcal{R}_y, \mathcal{R}_z)^T$, $c = \cos\alpha$, and $s = \sin\alpha$. The Cartesian components of $\boldsymbol{\mathcal{R}}$ are defined as

$$\mathcal{R}_x = \sqrt{1-\theta^2}\cos\varphi\ ,\ \mathcal{R}_y\sqrt{1-\theta^2}\sin\varphi\ ,\ \mathcal{R}_z = \theta, \tag{85}$$

where φ and θ are uncorrelated random numbers, which are taken from uniform distributions in the intervals $[0, 2\pi]$ and $[-1, 1]$, respectively.

On the other hand, a vector rotation can be performed [46]. The vector $\Delta\boldsymbol{v}_i = \boldsymbol{v}_i - \boldsymbol{v}_{cm} = \Delta\boldsymbol{v}_{i,\parallel} + \Delta\boldsymbol{v}_{i,\perp}$ is given by the component $\Delta\boldsymbol{v}_{i,\parallel} = (\Delta\boldsymbol{v}_i\boldsymbol{\mathcal{R}})\boldsymbol{\mathcal{R}}$ parallel to $\boldsymbol{\mathcal{R}}$ and $\Delta\boldsymbol{v}_{i,\perp} = \Delta\boldsymbol{v}_i - \Delta\boldsymbol{v}_{i,\parallel}$ perpendicular to $\boldsymbol{\mathcal{R}}$ (cf. Fig. 3 with $\Delta\boldsymbol{v}_i \equiv \boldsymbol{u}$). Rotation by an angle α transforms $\Delta\boldsymbol{v}_i$ into $\Delta\boldsymbol{v}_i' = \Delta\boldsymbol{v}_{i,\parallel} + \Delta\boldsymbol{v}_{i,\perp}'$. $\Delta\boldsymbol{v}_{i,\perp}'$ can be expressed by the vector $\Delta\boldsymbol{v}_{i,\perp}$ and the vector $\boldsymbol{\mathcal{R}} \times \Delta\boldsymbol{v}_{i,\perp}$, which yields

$$\begin{aligned} \boldsymbol{v}_i(t+h) &= \boldsymbol{v}_{cm}(t) + \cos\alpha\Delta\boldsymbol{v}_{i,\perp} + \sin\alpha\,(\boldsymbol{\mathcal{R}} \times \Delta\boldsymbol{v}_{i,\perp}) + \Delta\boldsymbol{v}_{i,\parallel} \\ &= \boldsymbol{v}_{cm}(t) + \cos\alpha\,[\Delta\boldsymbol{v}_i - (\Delta\boldsymbol{v}_i\boldsymbol{\mathcal{R}})\boldsymbol{\mathcal{R}}] \\ &\quad + \sin\alpha\,\boldsymbol{\mathcal{R}} \times [\Delta\boldsymbol{v}_i - (\Delta\boldsymbol{v}_i\boldsymbol{\mathcal{R}})\boldsymbol{\mathcal{R}}] + (\Delta\boldsymbol{v}_i\boldsymbol{\mathcal{R}})\boldsymbol{\mathcal{R}}\ , \end{aligned} \tag{86}$$

since the vector $\boldsymbol{\mathcal{R}} \times \Delta\boldsymbol{v}_{i,\perp}$ is perpendicular to $\boldsymbol{\mathcal{R}}$ and $\Delta\boldsymbol{v}_{i,\perp}$.

References

[1] H. Löwen, J. Phys.: Condens. Matter **13**, R415 (2001).

[2] G. Gompper, T. Ihle, D. M. Kroll, and R. G. Winkler, Adv. Polym. Sci. **221**, 1 (2009).

[3] M. Doi, J. Phys.: Condens. Matter **23**, 284118 (2011).

[4] P. J. Hoogerbrugge and J. M. V. A. Koelman, Europhys. Lett. **19**, 155 (1992).

[5] P. Español, Phys. Rev. E **52**, 1734 (1995).

[6] P. Español and P. B. Warren, Europhys. Lett. **30**, 191 (1995).

[7] G. McNamara and G. Zanetti, Phys. Rev. Lett. **61**, 2332 (1988).

[8] X. Shan and H. Chen, Phys. Rev. E **47**, 1815 (1993).

[9] X. He and L.-S. Luo, Phys. Rev. E **56**, 6811 (1997).

[10] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Oxford University Press, Oxford, 1994).

[11] F. J. Alexander and A. L. Garcia, Comp. in Phys. **11**, 588 (1997).

[12] A. L. Garcia, *Numerical Methods for Physics* (Prentice Hall, 2000).

[13] A. Malevanets and R. Kapral, J. Chem. Phys. **110**, 8605 (1999).

[14] A. Malevanets and R. Kapral, J. Chem. Phys. **112**, 7260 (2000).

[15] S. O. Nielsen, R. E. Bulo, P. B. Moore, and B. Ensing, Phys. Chem. Chem. Phys. **12**, 12401 (2010).

[16] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon Press, London, 1959).

[17] J. P. Boon and S. Yip, *Molecular Hydrodynamics* (Dover, New York, 1980).

[18] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, London, 1986).

[19] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1989).

[20] J. K. G. Dhont, *An Introduction to Dynamics of Colloids* (Elsevier, Amsterdam, 1996).

[21] M. Theers and R. G. Winkler, Soft Matter **10**, 5894 (2014).

[22] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).

[23] E. M. Purcell, Am. J. Phys. **45**(1), 3 (1977).

[24] C.-C. Huang, G. Gompper, and R. G. Winkler, Phys. Rev. E **86**, 056711 (2012).

[25] B. J. Alder and T. E. Wainwright, Phys. Rev. A **1**, 18 (1970).

[26] R. Zwanzig and M. Bixon, Phys. Rev. A **2**, 2005 (1970).

[27] M. H. Ernst, E. H. Hauge, and J. M. J. van Leeuwen, Phys. Rev. A **4**, 2055 (1971).

[28] E. Hauge and A. Martin-Löf, Journal of Statistical Physics **7**, 259 (1973).

[29] E. J. Hinch, J. Fluid Mech. **72**, 499 (1975).

[30] G. L. Paul and P. N. Pusey, J. Phys. A **14**, 3301 (1981).

[31] B. U. Felderhof, J. Chem. Phys. **123**, 044902 (2005).

[32] C. C. Huang, G. Gompper, and R. G. Winkler, J. Chem. Phys. **138**, 144902 (2013).

[33] M. Theers and R. G. Winkler, Phys. Rev. E **88**, 023012 (2013).

[34] P. Résibois and M. de Leener, *Classical Kinetic Theory of Fluids* (John Wiley & Sons, 1977).

[35] C. Cercignani, *The Boltzmann Equation* (Springer New York, 1988), vol. 67 of *Applied Mathematical Sciences*, p. 40.

[36] U. D. Schiller, *Thermal fluctuations and boundary conditions in the lattice Boltzmann method* (Doctoral Thesis, Johannes Gutenberg-Universität, Mainz, 2008).

[37] P. Bhatnagar, E. Gross, and M. Krook, Phys. Rev. **94**, 511 (1954).

[38] S. Chen and G. D. Doolen, Annu. Rev. Fluid Mech. **30**(1), 329 (1998).

[39] D. A. Fedosov, I. V. Pivkin, and G. E. Karniadakis, J. Comput. Phys. **227**, 2540 (2008).

[40] E. Moeendarbary, T. Y. NG, and M. Zangeneh, Int. J. Appl. Mechanics **1**, 737 (2009).

[41] M. Ripoll, in *Computational Condensed Matter Physics*, edited by S. Blügel, G. Gompper, E. Koch, H. Müller-Krumbhaar, R. Spatschek, and R. G. Winkler (Forschungszentrum Jülich, 2006), vol. 32 of *Matter and Materials*, chap. B5.

[42] M. P. Allen and F. Schmid, Mol. Simul. **33**, 21 (2007).

[43] X. Fan, N. Phan-Thien, S. Chen, X. Wu, and T. Y. Ng, Phys. Fluids **18**, 063102 (2006).

[44] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).

[45] R. Kapral, Adv. Chem. Phys. **140**, 89 (2008).

[46] E. Allahyarov and G. Gompper, Phys. Rev. E **66**, 036702 (2002).

[47] H. Noguchi, N. Kikuchi, and G. Gompper, EPL **78**, 10005 (2007).

[48] H. Noguchi and G. Gompper, Phys. Rev. E **78**, 016706 (2008).

[49] T. Ihle and D. M. Kroll, Phys. Rev. E **63**, 020201(R) (2001).

[50] T. Ihle and D. M. Kroll, Phys. Rev. E **67**, 066705 (2003).

[51] C.-C. Huang, A. Chatterji, G. Sutmann, G. Gompper, and R. G. Winkler, J. Comput. Phys. **229**, 168 (2010).

[52] E. Westphal, S. P. Singh, C.-C. Huang, G. Gompper, and R. G. Winkler, Comput. Phys. Comm. **185**, 495 (2014).

[53] E. Tüzel, T. Ihle, and D. M. Kroll, Phys. Rev. E **74**, 056702 (2006).

[54] T. Ihle and D. M. Kroll, Phys. Rev. E **67**, 066706 (2003).

[55] T. Ihle, E. Tüzel, and D. M. Kroll, Phys. Rev. E **70**, 035701 (2004).

[56] T. Ihle, E. Tüzel, and D. M. Kroll, Phys. Rev. E **72**, 046707 (2005).

[57] R. G. Winkler and C.-C. Huang, J. Chem. Phys. **130**, 074907 (2009).

[58] N. Kikuchi, C. M. Pooley, J. F. Ryder, and J. M. Yeomans, J. Chem. Phys. **119**, 6388 (2003).

[59] C. M. Pooley and J. M. Yeomans, J. Phys. Chem. B **109**, 6505 (2005).

[60] R. G. Winkler, in *Hierarchical Methods for Dynamics in Complex Molecular Systems*, edited by J. Grotendorst, G. Sutmann, G. Gompper, and D. Marx (Forschungszentrum Jülich GmbH, Jülich, 2012), vol. 10 of *IAS Series*.

[61] C.-C. Huang, A. Varghese, G. Gompper, and R. G. Winkler, Phys. Rev. E **accepted for publication** (2015).

[62] M. Ripoll, K. Mussawisade, R. G. Winkler, and G. Gompper, Phys. Rev. E **72**, 016701 (2005).

[63] A. Malevanets and J. M. Yeomans, Europhys. Lett. **52**, 231 (2000).

[64] M. Ripoll, K. Mussawisade, R. G. Winkler, and G. Gompper, Europhys. Lett. **68**, 106 (2004).

[65] K. Mussawisade, M. Ripoll, R. G. Winkler, and G. Gompper, J. Chem. Phys. **123**, 144905 (2005).

[66] S. Poblete, A. Wysocki, G. Gompper, and R. G. Winkler, Physical Review E **90**, 033314 (2014).

[67] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982).

[68] S. H. Lee and R. Kapral, J. Chem. Phys. **121**, 11163 (2004).

[69] A. Lamura, G. Gompper, T. Ihle, and D. M. Kroll, Europhys. Lett. **56**, 319 (2001).

[70] Y. Inoue, Y. Chen, and H. Ohashi, J. Stat. Phys. **107**, 85 (2002).

[71] I. O. Götze, H. Noguchi, and G. Gompper, Phys. Rev. E **76**, 046705 (2007).

[72] J. K. Whitmer and E. Luijten, J. Phys.: Condens. Matter **22**, 104106 (2010).

[73] M. Yang and M. Ripoll, Soft Matter **10**, 1006 (2014).

[74] A. W. Lees and S. F. Edwards, J. Phys. C **5**, 1921 (1972).

[75] L. Cannavacciuolo, R. G. Winkler, and G. Gompper, EPL **83**, 34007 (2008).

[76] H. Noguchi and G. Gompper, Proc. Natl. Acad. Sci. USA **102**, 14159 (2005).

[77] J. L. Mcwhirter, H. Noguchi, and G. Gompper, Proc. Natl. Acad. Sci. USA **106**, 6039 (2009).

[78] R. Chelakkot, R. G. Winkler, and G. Gompper, EPL **91**, 14001 (2010).

[79] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Science* (Cambridge University Press, 2002).

[80] S. H. Strogatz, *Sync: How Order Emerges From Chaos In the Universe, Nature, and Daily Life* (Hyperion, 2004).

[81] R. R. Bennett and R. Golestanian, Phys. Rev. Lett. **110**, 148102 (2013).

[82] K. Drescher, R. E. Goldstein, N. Michel, M. Polin, and I. Tuval, Phys. Rev. Lett. **105**, 168101 (2010).

[83] R. E. Goldstein, M. Polin, and I. Tuval, Phys. Rev. Lett. **107**, 148103 (2011).

[84] J. S. Guasto, K. A. Johnson, and J. P. Gollub, Phys. Rev. Lett. **105**, 168102 (2010).

[85] E. Lauga and R. E. Goldstein, Phys. Today **65**, 30 (2012).

[86] B. Qian, H. Jiang, D. A. Gagnon, K. S. Breuer, and T. R. Powers, Phys. Rev. E **80**, 061919 (2009).

[87] M. Kim and T. R. Powers, Phys. Rev. E **69**, 061910 (2004).

[88] M. J. Kim, J. C. Bird, A. J. Van Parys, K. S. Breuer, and T. R. Powers, Proc. Natl. Acad. Sci. USA **100**, 15481 (2003).

[89] M. Reichert and H. Stark, Eur. Phys. J. E **17**, 493 (2005).

[90] S. Y. Reigh, R. G. Winkler, and G. Gompper, Soft Matter **8**, 4363 (2012).

[91] S. Y. Reigh, R. G. Winkler, and G. Gompper, PLoS ONE **8**, e70868 (2013).

[92] E. W. Knight-Jones, Q. J. Microsc. Sci. **95**, 503 (1954).

[93] D. R. Brumley, M. Polin, T. J. Pedley, and R. E. Goldstein, Phys. Rev. Lett. **109**, 268102 (2012).

[94] M. A. Sleigh, *The Biology of Cilia and Flagella* (Pergamon Press, Oxford, 1962).

[95] M. Polin, I. Tuval, K. Drescher, J. P. Gollub, and R. E. Goldstein, Science **325**(5939), 487 (2009).

[96] R. Stocker and W. M. Durham, Science **325**(5939), 400 (2009).

[97] B. Afzelius, Science **193**(4250), 317 (1976).

[98] Q. Wang, J. Pan, and W. J. Snell, Cell **125**, 549 (2006).

[99] J. H. E. Cartwright, O. Piro, and I. Tuval, Proc. Natl. Acad. Sci. USA **101**(19), 7234 (2004).

[100] P. Lenz and A. Ryskin, Phys. Biol. **3**, 285 (2006).

B 4 Colloids and their interactions

P. R. Lang, S. De Sio
Institute of Complex Systems / Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 van der Waals Interaction **3**
2.1 van der Waals interaction between atoms and molecules 3
2.2 van der Waals interactions between colloidal spheres 4
2.2.1 The Derjaguin Approximation 7

3 Electrostatic potentials and potential energies in the Debye-Hückel approximation **10**
3.1 Point charges 10
3.2 Charged spheres 12
3.3 Electrolyte concentration 14

4 The DLVO Theory and Colloidal Stability **14**

5 Non-DLVO interactions **17**
5.1 Depletion interaction 17
5.1.1 Non-spherical depletants 19
5.2 Steric repulsion 21

6 Experimental Methods **22**
6.1 Surface forces apparatus 23
6.2 Total internal reflection microscopy 24
6.2.1 Depletion interactions measured with TIRM 27

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

In 1861 the Scottish Chemist Thomas Graham published results of a study he had conducted with very simple means. Aqueous solutions of various materials were placed in glass cylindrical with a bottom made of parchment or any kind of animal membrane like pig bladder, which was placed into a basin of pure water. He realized that some of the materials, like salts would easily diffuse cross the membrane into the water reservoir, while others, like starch, gum arabic or gelatin did not pass through the membrane. The first kind of substances was known to crystalize while the second one formed sticky amorphous layers on the membranes. Accordingly Graham distinguished between crystalloids and colloids, where the latter term derives from the Greek $\kappa o\lambda\lambda\alpha$ (glue) and $\varepsilon\iota\delta o\varsigma$ (shape, appearance). Probably this was the reason why colloids for a long time had the reputation of "things indefinite in shape, indefinite in chemical composition and physical properties, fickle in chemical department, things infilterable and generally unmanageable", as Ernest S. Hedges put it in the preface of his book in 1931 [1]. Nowadays, after the ground breaking work of van't Hoff, Zsigmondy, Ostwald, Staudinger and many others, we know that the class of substances, which Graham called colloids has to be subdivided in a number of different systems spanning the range from metallic nano-particles over synthetic polymers to biological macromolecules like proteins and DNA. A broader definition of colloidal systems includes also micellar surfactant solutions, emulsions, foams and aerosols.

A general common feature of all these systems is their typical length scale, which is the basis the nowadays effective IUPAC definition[2] of the term colloidal:"State of subdivision such that the molecules or polymolecular particles dispersed in a medium have at least one dimension between approximately 1 nm and 1 μm, or that in a system discontinuities are found at distances of that order." Thus, the typical length scale of a colloidal particle is by at least one order of magnitude larger than the molecules of the solvent suspending it. Another important feature of colloids is that they are undergoing Brownian motion in liquid suspension at typical timescales approximately ranging from 10 microseconds to 100 milliseconds, which is six to ten orders of magnitude slower then typical relaxation times of molecular liquids. The dynamics of colloidal particles in suspension will be subject of chapter B6 by Jan Dhont.

Finally colloidal systems can also be characterized by their typical pair interaction energies which is of the order of a few thermal units, k_BT at equilibrium separation distance. The appearingly simplest contribution to the pair interaction potentials is the so called hard body interaction. Given their large size, it is immediately obvious that colloidal particles will interact via excluded volume when suspended in a solvent. This leads to a series of phase transitions when the particle volume fraction is increased. In a suspension of hard spheres, i.e. spherical particles of radius R which have an interaction potential that is zero if the particles are at distances $r > 2R$ and that is infinitely repulsive at $r \leq 2R$, at a sphere volume fraction of about 50% a transition is observed from a liquid sate to a state which shows characteristics of a crystal. For colloids with long ranging interactions all the pase transitions and co-existences known from real gases and some more, like the formation of different types of glassy states, can be observed, depending on the details of type and range of interactions.

In this chapter we will discuss the most important types of pair interactions between colloids in suspension, which have longer range than hard body interactions. For the sake of brevity, we will restrict the discussion to particles of spherical shape bearing in mind, that the basic concepts can be transferred to differently shaped colloids like cylinders and discs by applying Derjaguin–approximation [5]. First we will describe van der Waals interactions, which are generally attractive, and electrostatic interactions, which are repulsive in most cases and very

often take the role of counterbalancing the attractive forces and to make colloidal suspensions stable against aggregation precipitation. The superposition of these two types of interaction form the basis of the celebrated DLVO-theory [3, 4], which is named after the pioneers of colloidal science, Derjaguin Landayu, Verwey and Overbeek. Later, two types of non-DLVO interactions will be discussed, the depletion interaction which is attractive and repulsive steric interactions. We will briefly describe two experimental techniques which are used to measure pair interaction potentials and finally we will discuss the results of a recent study experimental study.

2 van der Waals Interaction

2.1 van der Waals interaction between atoms and molecules

The origin of van der Waals interactions between colloidal particles is the interaction between the atoms or molecules of which they consist. We will therefore start with a heuristic discussion of the von der Waals interactions between uncharged atomic and molecular entities in the gas phase. These may be very comprehensively explained as dipolar interactions, which consist of an orientational contribution, an inductive contribution and the so called dispersion forces [6]. The orientational contribution, which is also termed Keesom interaction, is the change of free energy associated with the approach of two molecular dipoles from infinite separation to a distance r accompanied with a potential weighted orientation distribution of the dipole moments m. Quantitatively it is given by

$$w_o(r) = -\frac{m_1^2 m_2^2}{3k_B T(4\pi\varepsilon_0)^2}\frac{1}{r^6}. \tag{1}$$

where ε_0 is the vacuum permittivity. Additionally to their direct interaction, which is described by eq. 1 permanent dipoles will induce additional dipole moments in other molecules if those are polarisable. This causes a further contribution to the interaction energy, sometimes called Debye interaction, which is proportional to the permanent moments squared of the interacting molecules, like the orientation interaction. Further the induced dipolar interaction is linearly proportional to the respective polarizabilities α_j. The final expression for the induced dipolar interaction between two different molecules is given by

$$w_i(r) = -\frac{m_1^2\alpha_2 + m_2^2\alpha_1}{(4\pi\varepsilon_0)^2}\frac{1}{r^6}. \tag{2}$$

Note that the induced dipolar interaction decays with $1/r^6$ like the orientation interaction.
The types of interaction described so far are basically electrostatic interactions, which require that at least one of the particles taking part carries a permanent dipole moment. There is however a third contribution, which acts between all atoms and molecules even if they are completely non–polar like noble gas atoms: the so–called London dispersion energy. In an intuitive picture this can be understood on the basis of Bohr's atomic model, as the interaction between fluctuating dipoles which come about by the orbiting of electrons around the nucleus. It is related to the atom's or molecules's ionization energy I by

$$w_d(r) = -\frac{4\alpha^2 I}{(4\pi\varepsilon_0)^2 r^6}. \tag{3}$$

Table 1: *Numerical values for the orientation contribution C_o, the contribution of poalrizability C_i and the dispersion contribution C_d to van der Waals interactions of like molecules.*

	$\alpha/(4\pi\varepsilon_0)$	m	I	C_o	C_i	C_d	C_{vdW}	C_d/C_{vdW}
	10^{-30} m	Cm	10^{-18} J	10^{-79} Jm^6				
HCl	2.36	3.6	1.35	11	6	106	123	0.86
CH_3Cl	4.56	6.24	1.2	101	32	282	415	0.68
NH_3	2.26	4.9	1.08	38	10	63	111	0.57
H_2O	1.48	6.17	1.34	96	10	33	139	0.24

Apart from a factor this is equal to the expression London derived on the basis of quantum mechanical perturbation calculations [7, 8, 9]. For two different atoms London's expression is

$$w_d(r) = -\frac{3}{2}\frac{\alpha_1\alpha_2}{(4\pi\varepsilon_0)^2 r^6}\frac{I_1 I_2}{I_1 + I_2} \tag{4}$$

which simplifies for two identical atoms to

$$w_d(r) = -\frac{3}{4}\frac{\alpha^2 I}{(4\pi\varepsilon_0)^2 r^6} \tag{5}$$

It is important to note that dispersion energies scale with $1/r^6$ like orientation interactions and induced dipolar interactions. Thus, there are three types of interaction contributing to the total interaction energy between two atoms or molecules, which all scale with $1/r^6$. The sum of these contributions is called the van der Waals interaction energy

$$w_{vdW}(r) = -\frac{1}{r^6}\left(C_o + C_i + C_d\right) \equiv -C_{vdW}\frac{1}{r^6} \tag{6}$$

where the parameters C_o, C_i and C_d are defined by eqs. 1,2 and 4 respectively. Numerical values for these parameters are listed in Table 1, which show a maybe unexpected general trend

$$C_d \gg C_o > C_i. \tag{7}$$

In most cases the dispersion forces are the dominating contribution, as is displayed in the distance profile of the van der Waals interaction shown in Fig. 1 for hydrogen chloride. Only in cases where the permanent molecular dipoles are very large and the polarizability is very low, like in the case of water, the contribution of the dispersion forces is smaller than fifty percent. In these cases the orientational contribution will be the largest one, as shown for the case of water molecules in the right panel of in Fig. 1.

2.2 van der Waals interactions between colloidal spheres

In the preceding sections we discussed the attractive interactions between atoms or molecules, which are the constituents of colloidal particles. The forces and energies acting between colloidal particles are of the same nature and in a first approach to calculate the interaction between colloidal particles we assume simple pair wise additivity of the molecular interaction potentials. First we will calculate the potential profile $W_{MS}(r)$ between a single molecular entity and

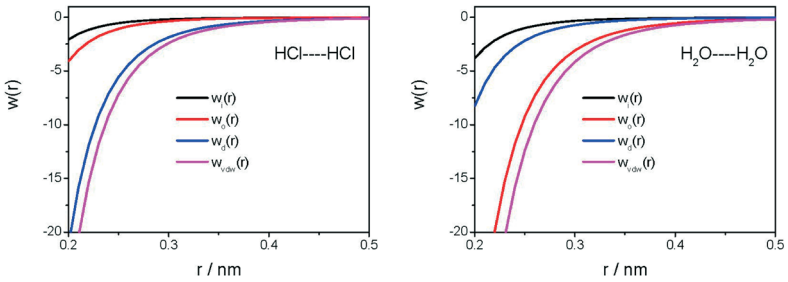


Fig. 1: *Distance profiles for the orientation contribution w_o, the contribution of polarizability w_i and the dispersion contribution w_d to van der Waals interactions w_{vdW} between two HCl–molecules (left) and between two H_2O–molecules (right.). The energy scale is in units of k_BT*

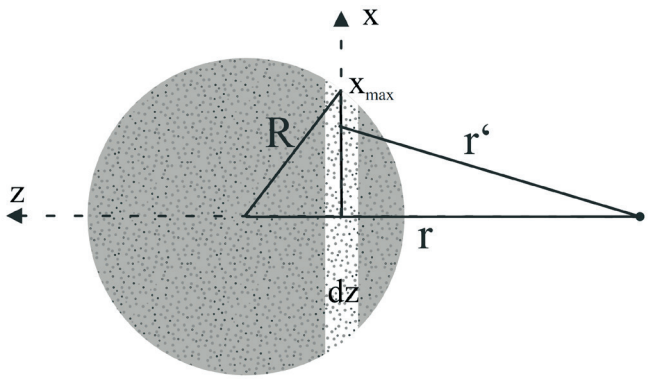


Fig. 2: *Sketch illustrating the integration over the interactions of a single molecule with all molecules constituting a colloidal sphere.*

a colloidal sphere consisting of N molecules of the same kind, which is the sum over all N molecule–molecule interactions

$$W_{MS}(r) = -\sum_{j=1}^{N} \frac{C_{vdw}}{r_j'^6} \tag{8}$$

where r is the distance of the molecule from the center of the sphere and r'_j represents the separation between the molecule outside the sphere and the j_{th} molecule within the sphere, as depicted in Fig 2. From here on we will use a capital W for the interaction energy between a molecular entity and a large body, to distinguish it from the interactions between to atomic or molecular objects. Replacing the sum by an integral we obtain

$$W_{MS}(r) = -\rho C_{vdw} \int \frac{1}{r'^6} dV \tag{9}$$

where $\rho = dN/dV$ is the number density of molecules within the sphere.

By replacing the volume element by $dV = 2\pi x dx dz$, noticing that $r' = \sqrt{x^2 + z^2}$ and realizing that the maximum value the x–coordinate can take, depends on z by $x_{max} = \sqrt{R^2 - (r-z)^2}$ we obtain

$$\begin{aligned} W_{MS}(r) &= -2\pi\rho C_{vdw} \int_{r-R}^{r+R} dz \int_0^{\sqrt{R^2-(r-z)^2}} dx \frac{x}{(x^2+z^2)^3} \\ &= -2\pi\rho C_{vdw} \int_{r-R}^{r+R} dz \frac{1}{4}\left[\frac{1}{z^4} - \frac{1}{(R^2 - r^2 + 2rz)^2}\right] \\ &= -\frac{4\pi\rho C_{vdw}}{3} \frac{R^3}{(r-R)^3(r+R)^3} \end{aligned} \tag{10}$$

If we now regard the molecule as being part of a second sphere a similar, though more tedious integration leads to the van der Waals interaction potential between two spheres of the same chemical composition with the radii R_1 and R_2

$$\phi_{SS}^{vdW}(r) = -\frac{A_H}{6}\left(\frac{2R_1R_2}{r^2-(R_1+R_2)^2} + \frac{2R_1R_2}{r^2-(R_1-R_2)^2} + ln\left[\frac{r^2-(R_1+R_2)^2}{r^2-(R_1-R_2)^2}\right]\right). \tag{11}$$

Here we use the symbol ϕ for the potential profile between two colloidal bodies, as we will do throughout the remainder of the text. Further, we have introduced a new pre–factor, which is named in the honor of Hugo Christian Hamaker, $A_H = \pi^2\rho^2 C_{vdW}$.
It is instructive investigate eq. 11 for the limiting cases of of very large spheres where $R/(r - 2R) \to \infty$ and very large distances where $2R/r \to 0$. For the first case we obtain

$$\lim_{R/(r-2R)\to\infty} = -\frac{A_H R}{12(r-2R)} \tag{12}$$

showing that for spheres which are large compared to their separation distance, the van der Waals attraction decays very slowly, i.e. linearly with reciprocal distance. For very large distances

$$\lim_{2R/r\to 0} = -\frac{16A_H R}{9}\left(\frac{R}{r}\right)^6, \tag{13}$$

yielding again is the distance dependence known from atomic or molecular objects.
In deriving eq. 11 we assumed pairwise additivity of the dipole–dipole interactions. This is of course a very strong restriction, for example it neglects the influence of a third atom on the induced dipole contribution and dispersion interaction between an other pair of atoms. This may be a valid approximation in highly dilute systems, i. e. gases, but not in condensed materials. Further this approximation can not be readily extended to the situation where two large particles are interacting through a dielectric medium. e. g. a solvent. These problems are avoided in the Lifshitz theory [10] where the particles are treated as a dielectric continuum neglecting their atomic structure. In this theory the forces are calculated on the basis of the particles' dielectric bulk properties. Fortunately we do not have to deal with the Lifshitz theory in detail, because the analytic form of eq. 11 remains the same. The only thing that changes is the Hamaker constant which becomes a function of the dielectric constants of the media involved. For two particles of refractive index n_1 and n_2 interacting across a dielectric medium with n_3 and the corresponding zero frequency dielectric constants ε_i this is given approximately as

$$A_H \approx \frac{3}{4}k_BT\left(\frac{\varepsilon_1-\varepsilon_3}{\varepsilon_1+\varepsilon_3}\right)\left(\frac{\varepsilon_2-\varepsilon_3}{\varepsilon_2+\varepsilon_3}\right) + \frac{3h\nu_e}{8\sqrt{2}} \frac{(n_1^2-n_3^2)(n_2^2-n_3^2)}{(n_1^2+n_3^2)^{1/2}(n_2^2+n_3^2)^{1/2}\left[(n_1^2+n_3^2)^{1/2}+(n_2^2+n_3^2)^{1/2}\right]} \tag{14}$$

Table 2: *Limiting cases for eq. 14*

	A_H	ϕ^{vdW}
$n_1 \neq n_2$,$n_3 = 1$	> 0	< 0
$n_1 = n_2 \neq n_3$	> 0	< 0
$n_1 < n_3 < n_2$	< 0	> 0

For the simple case of two particles consisting of the same material with refractive index n_1 this simplifies to

$$A_H \approx \frac{3}{4} k_B T \left(\frac{\varepsilon_1 - \varepsilon_3}{\varepsilon_1 + \varepsilon_3} \right)^2 + \frac{3h\nu_e}{16\sqrt{2}} \frac{\left(n_1^2 - n_3^2\right)^2}{\left(n_1^2 + n_3^2\right)^{3/2}}. \tag{15}$$

where $h = 6.63^{-34}$Js is Planck's constant and ν_e is the main electronic absorption frequency of the material, which is typically of the order 10^{14}Hz. Together with the values for refractive index and dielectric constants of typical colloidal material and solvents the Hamaker constant can be calculated for most materials to be of the order of a fraction of $k_B T$ up to several thermal units.

It is important to note that this equation shows the route to very effectively tune the van der Waals interaction between colloidal spheres. In situations where the refractive index of a material is approximately the square of its dielectric constant, e. g. for most organic materials, the Hamaker constant may be made approximately zero, if the colloids are suspended in a solvent which matches their refractive index. On the other hand, for the case of water as a solvent, A_H will never become smaller than about $0.7 k_B T$, even if the particles had the exactly the same refractive index as water (which will never happen by the way) due to the very large zero frequency dielectric constant of $\varepsilon_{H_2O} = 81$.

Further, it is very instructive to have a look at certain limiting cases of eq. 14. As can be seen from Table 2 the potential between two bodies interacting through vacuum is always attractive, as is the potential between two bodies of the same material interacting across any medium. Repulsive van der Waals interactions are found only between particles with different refractive index n_1 and n_2 interacting through a medium, the refractive index of which is intermediate between n_1 and n_2. This is for example the case when a vapor phase $n_1 \approx 1$ interacts across a liquid phase $1.3 \lesssim n_3 \lesssim 1.5$ with glass $n_2 \gtrsim 1.5$. The repulsive van der Waals potential between the vapor/liquid and the liquid/glass interface is the reason why most liquids tend to form a positive meniscus at the glass wall of a beaker or a test tube.

2.2.1 The Derjaguin Approximation

In the last section, we have introduced an exact description for the distance dependence of van der Waals interactions between two colloidal spheres. In many cases it is however convenient to use a less rigorous approximate approach, which is based on the calculation of the interaction potential between two flat interfaces, from which the forces between curved interfaces are derived. This approximation, named after Boris W. Derjaguin, will be outlined qualitatively in the following.

Similar to the procedure, which was applied to calculate the interaction energy between a molecule and a colloidal sphere the interaction between a molecule and a flat solid wall W_{MW}

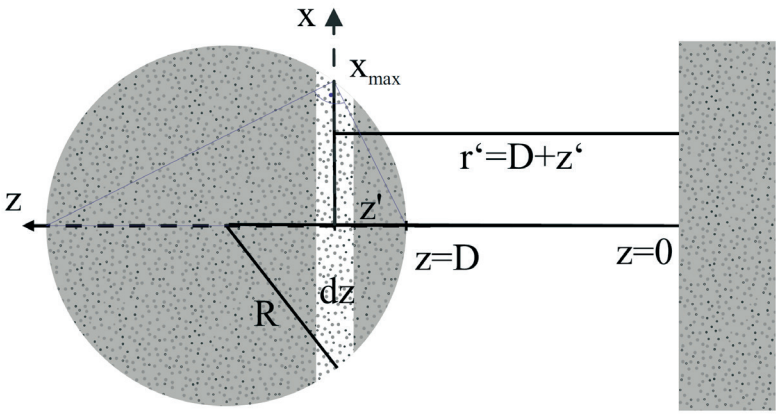


Fig. 3: *Sketch illustrating the integration over the interactions of a colloidal sphere with a flat wall.*

can be obtained as

$$W_{MW}(D) = -\frac{\pi\rho C_{vdw}}{6}\frac{1}{D^3}. \tag{16}$$

where D is the shortest distance between the molecule and the wall surface. Along the same line of arguments, the van der Waals interaction between two flat walls can be calculated to be

$$\phi_{WW}^{vdW,a}(D) = -\frac{\pi\rho^2 C_{vdW}}{12D^2}. \tag{17}$$

Here the superscript ^{a} indicates an energy per unit area, which has to be introduced, because integration over all molecules of an infinitely extended wall would result diverging potential values.
In the next step we are going to calculate the potential between a wall and a spherical colloid, for which we have to integrate over all molecules within the sphere, interacting with the wall according to eq. 16. According to Fig. 3 all molecules within an infinitely thin slab of volume $dV = 2\pi x_{max}^2 dz$ at $z = r - R + z'$ have the same distance $r' = D + z'$ from the wall. The radius of this slice, x_{max}, depends on it's z–position like $x_{max}^2 = (2R - z')z'$, as is immediately seen considering the blue triangle in Fig.3. With this, the interaction between the sphere and the (infinitely extended) wall is now

$$\begin{aligned}\phi_{SW}^{vdw}(D) &= \int dV \rho W_{MW}(D + z') \\ &= -\frac{\pi^2\rho^2 C_{vdW}}{6}\int_0^{2R} dz' \frac{(2R - z')z'}{(D + z')^3}\end{aligned} \tag{18}$$

because the Jacobian $dz/dz' = 1$.
Since the molecule–wall interaction drops off with the inverse third power of distance we may assume that only molecules at $z' << 2R$ contribute significantly to the sphere–wall interaction.

In this case the numerator in the integrand of eq. 18 is $(2R - z')z' \approx 2Rz'$ and the integral attains the general form

$$\int dx x^m (a+bx^n)^p = \frac{1}{a(m+1)} \left[x^{m+1}(a+bx^n)^{p+1} - (m+n+1+np)b \int dx x^{m+n}(a+b^n)^p \right].$$

Identifying $m = n = b = 1$, $p = -3$, $a = D$ and $x = z'$ we obtain

$$\int_0^{2R} dz' \frac{z'}{(D+z')^3} = \frac{1}{2D} \frac{z'^2}{(D+z')^2}|_0^{2R} \tag{19}$$

which reduces to $1/(2D)$ in the limit of $R >> D$. Finally the sphere-wall interaction becomes

$$\phi_{SW}^{vdw}(D) = -\frac{\pi^2 \rho^2 C R}{6D}. \tag{20}$$

Note that the interaction between a sphere and a wall decays only linearly with the reciprocal distance.
The van der Waals force between a sphere and a wall is according to eq. 20

$$F_{SW}(D) = -\frac{\partial \phi_{SW}^{vdw}(D)}{\partial D} = -\frac{\pi^2 \rho^2 C R}{6D^2} \tag{21}$$

Comparison of the r.h.s of eq. 21 with the expression for the interaction between two planar walls, eq. 17, of unit area shows that

$$F_{SW}(D) = 2\pi R \phi_{WW}^a \tag{22}$$

This important relation is the essence of the so-called Derjaguin approximation. It relates the force between curved surfaces to the interaction energy between two planar walls at the same distance. It holds for $D << R$ and ranges of interaction which are small compared to the particle radii. It is important to note that this approximation holds for any kind of interaction forces, which decrease with with separation distance. Further it can be generalized to the case of two curved surfaces as

$$F_{SS}(D) = -2\pi \frac{R_2 R_1}{R_1 + R_2} \phi_{WW}^a(D).$$

This allows us to calculate immediately the van der Waals force between two colloidal particles which for two spheres of same radius $R = R_1 = R_2$ and composition is

$$F_{SS}^{vdw}(D) = 2\pi \frac{R}{2} \frac{\pi \rho^2 C_{vdW}}{12D^2} \tag{23}$$

which can be easily integrated to obtain the potential profile as

$$\begin{aligned} \phi_{SS}^{vdw}(D) &= -\int_D^\infty dD F_{SS}^{vdw}(D) \\ &= -\frac{A_H R}{12D}. \end{aligned} \tag{24}$$

This is identical with the limiting case of the exact expression given by of eq. 12 for separation distances which are small compared to the particle radii

3 Electrostatic potentials and potential energies in the Debye-Hückel approximation

In the preceeidng section we have seen that van der Waals interactions between colloidal particles across a dielectric medium are always attractive. This implies that all colloidal particles have a strong tendency to aggregate in any suspending liquid. Since a large fraction of biologically active molecules, like protein or DNA are of colloidal size, this would have fatal consequences, if there was not a least one other contribution to interaction, which counter-balances van der Waals attraction. In most of the cases it is electrostatic repulsion, which stabilizes colloids against aggregation In this section we will give a heuristic derivation of the mean electrostatic potential involved in electrostatic interactions among charged macromolecules dissolved in a polar solvent. For a thorough theoretical treatment, the reader is referred to references [11] and [12] and work cited therein.

3.1 Point charges

The mean potential emerging from a charged object cannot be simply described by a Coulomb interaction, because the electrostatic force between two charged objects will be screened by the micro–ions which are inevitably present in the solution. In the simplest case these are the dissociated counter-ions, but in many practical situations, e. g. in biological systems at physiological conditions there are additional micro-ions due to the presence of low molar mass electrolytes. To include this screening, a more involved model is required, which accounts, in a mean–field way, for the distribution of all charges in the solution.
The mean electric potential due to a point–like source, $u_P(\vec{r})$, at a given separation from the source is related to the charge density $\rho(\vec{r})$ at $\vec{r}$ by the Poisson equation (in SI-units)

$$\nabla^2 u_P(\vec{r}) = -\frac{\rho(\vec{r})}{\varepsilon\varepsilon_0}. \tag{25}$$

where ε is the dielectric constant of the solvent and ε_0 the vacuum permittivity. Since the potential emerging from a point source is spherically symmetric it is convenient to express the Laplace operator in polar coordinates omitting the angular dependence

$$\nabla^2 u_P(\vec{r}) = \frac{1}{r}\frac{\partial^2}{\partial r^2}\left(r u(r)\right) = -\frac{\rho(r)}{\varepsilon\varepsilon_0} \tag{26}$$

with $u_P(r)$ denoting the radially symmetric electrostatic potential at distance r from the point source. The mean charge density at r is the sum of the number densities $\mathcal{N}_j$ of the individual micro–ion species multiplied by their respective charge

$$\rho(r) = \sum_j \mathcal{N}_j(r) e Z_j. \tag{27}$$

Here e is the elementary charge and Z_j is the valency, i. e. the number of elementary charges of micro–ion species j. The number density of a given species at a given position is approximated by a Boltzmann–distribution

$$\mathcal{N}_j(r) = \mathcal{N}_j^0 \exp\left\{-\beta Z_j e u_P(r)\right\} \tag{28}$$

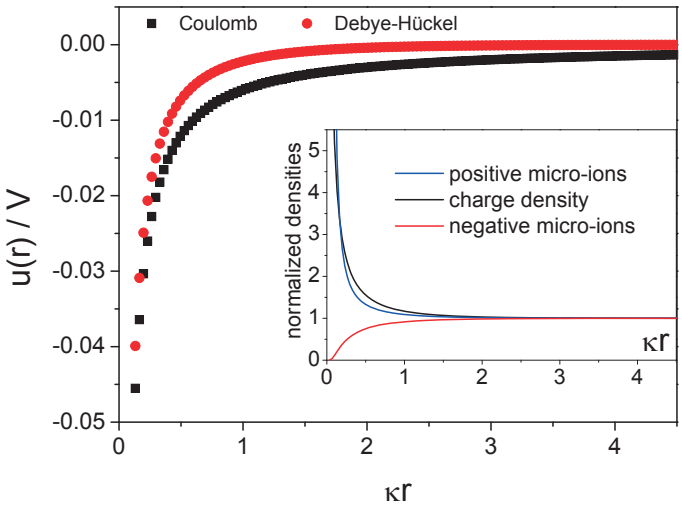


Fig. 4: *The mean electrostatic potential of a point source with one negative elementary charge in a 0.01 M NaCl solution (red bullets) and a hypothetical Coulomb potential in the same solution (black squares). Inset: Corresponding Densities of positive micro–ions (blue), negative micro–ions (red) and the total charge (black) around the source. The data are normalized to the respective values at infinite distance.*

where $\mathcal{N}_j^0$ are the number densities at zero potential, e. g. infinitely far away from the source and $\beta = 1/k_BT$ is the inverse thermal energy unit. Insertion of eq. 28 into eq. 25 gives the so–called Poisson–Boltzmann equation for spherical symmetry

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}\left(ru_P(r)\right) = -\frac{1}{\varepsilon\varepsilon_0}\sum_j \mathcal{N}_j^0 \exp\left\{-\beta Z_j e u_P(r)\right\} \tag{29}$$

which in general can not be solved analytically due to its non–linearity. However, for the limiting case of small potentials, i. e. $\beta Z_j e u_P(r) \ll 1$ the charge density can be approximated by a Taylor–expansion truncated after the linear term. With this we obtain the linearized Poisson–Boltzmann equation for spherical symmetry

$$\begin{aligned}\frac{1}{r}\frac{\partial^2}{\partial r^2}\left(ru_P(r)\right) &\approx \frac{\sum_j \mathcal{N}_j^0 e Z_j}{\varepsilon\varepsilon_0}\left(1-\left(-\beta Z_j e u_P(r)\right)\right) \\ &= \frac{\beta e^2 \sum_j \mathcal{N}_j^0 Z_j^2}{\varepsilon\varepsilon_0} u_P(r),\end{aligned} \tag{30}$$

often referred to as the Debye–Hückel approximation. Note that the first term of the series expansion vanishes due to the requirement that the system in total shall be electrically neutral. With the definition $\kappa^2 = \beta e^2 \sum_j \mathcal{N}_j^0 Z_j^2/\varepsilon\varepsilon_0$ we obtain the Debye–Hückel equation

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}\left(ru_P(r)\right) = \kappa^2 u_P(r) \tag{31}$$

whose general solution for spherical symmetry and $r > 0$ is

$$u_P(r) = \frac{A_1}{r} \exp\{\kappa r\} + \frac{A_2}{r} \exp\{-\kappa r\}, \tag{32}$$

which can be easily verified by insertion. The integration constants A_1 and A_2 are determined by the boundary conditions, the first of which requires that the potential must not diverge at large distances. This leads immediately to $A_1 = 0$. Second, the potential should be of Coulombic nature without screening, that is as $\sum_j \mathcal{N}_j \to 0$ and consequently $\kappa \to 0$. In this case $A_2 = q/\left(4\pi\varepsilon\varepsilon_0 r\right)$. Thus, we obtain the solution

$$u_P(r) = \frac{Z_P e}{4\pi\varepsilon\varepsilon_0} \frac{1}{r} \exp\left\{-\kappa r\right\} \tag{33}$$

for the mean electrostatic potential of point like source with a charge $q = Z_P e$ in the Debye–Hückel approximation.

The potential depends on the distance from the source in a complicated way. While it decays approximately as $1/r$ at small distances, it approaches an exponentially decaying behavior far from the source. The associated decay length, κ^{-1} is usually termed Debye screening length. To demonstrate the effect of screening, in Fig. 4 we plot the potential of a point source with one negative elementary charge as a function of normalized distance, κr. As an example, we compare the screened mean electrostatic potential in an aqueous 0.01 M NaCl–solution to a hypothetical bare Coulomb potential in the same solution. The magnitude of the Debye–Hückel potential is always smaller than that of the Coulomb potential due to the screening and the ratio of the two is $1/e$ at $\kappa r = 1$. In the inset of Fig. 4 the charge density distribution is displayed together with the number density distributions of the positive and the negative micro–ions. There is a region where the magnitude of the charge density and the number density of the oppositely (with respect to the source) charged micro–ions is significantly enhanced. In the same region the number density of the of the like charged micro–ions is substantially smaller than at infinite distance. This region is usually referred to as the diffuse (double) layer and κ^{-1} is used as an estimate for its thickness.

3.2 Charged spheres

An expression quite similar to eq. 33 can be derived for the mean electrostatic potential emerging from a uniformly charged sphere, $U_S(\vec{r})$ with radius R suspended in an electrolyte solution, if the diffuse layer thickness is sufficiently large as compared to the sphere radius, i. e. typically for $\kappa R \lesssim 5$. Now, the linearized Poisson–Boltzmann equation, which is valid for small surface potentials has to be solved for a sphere

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}\left(rU_S\left(r\right)\right) = \begin{cases} \kappa^2 U_S(r) & \text{for} \quad r > R \\ 0 & \text{for} \quad 0 < r < R \end{cases}. \tag{34}$$

The general solution of this linear differential equation outside the sphere is again given by eq. 32, whereas inside the sphere $U_S(r < R) = A_3/r + A_4$. The constant $A_3 = Z_S e/\varepsilon\varepsilon 0$, where Z_S is the number of elementary charges on the sphere, and A_2 and A_4 follow from the condition that $U_S(r)$ and $4\pi\varepsilon\varepsilon_0 \partial U_S(r)/\partial r$ have to be continuous for $r = R$ as

$$\begin{aligned} A_2 &= \frac{Z_S e}{4\pi\varepsilon\varepsilon_0} \frac{\exp\{\kappa R\}}{1+\kappa R} \\ A_4 &= \frac{Z_S e}{4\pi\varepsilon\varepsilon_0} \frac{\kappa}{1+\kappa R}. \end{aligned} \tag{35}$$

Thus, the mean electrostatic potential outside a sphere with Z_S uniformly distributed elementary charges in the Debye–Hückel approximation is given by

$$U_S(r) = \frac{Z_S e}{4\pi\varepsilon\varepsilon_0} \frac{exp\{\kappa R\}}{1+\kappa R} \frac{\exp\{-\kappa r\}}{r} \quad \text{for} \quad \mathrm{r} > \mathrm{R} \tag{36}$$

which reduces to eq. 33 for $R \to 0$. Note that eq. 34 is valid for all κr as long as $\beta\phi_S(r = R) \ll 1$ while the potential form of eq. 36 is a good approximation only for $\kappa R \lesssim 5$.
For large spheres with a comparatively thin diffuse layer, typically $\kappa R \gg 5$ a different approximation is more appropriate. In the limiting case of $R \to \infty$, i. e. a charged flat wall, with a constant surface charge density σ we have to solve the one dimensional form of the linearized Poisson–Boltzmann equation

$$\frac{\partial^2 U_W(D)}{dD^2} = \kappa^2 U_W(D) \tag{37}$$

where D is the distance from the wall. It can be easily verified by insertion that the solution of this differential equation is

$$U_W(D) = U_0 \exp\{-\kappa D\} \quad \text{for} \quad \mathrm{D} \geq 0 \tag{38}$$

with U_0 the potential at zero distance, i. e. at the surface of the wall. This surface potential is related to the charge density by the Grahame–equation, as

$$\sigma = \sqrt{4\varepsilon\varepsilon_0 k_B T \sum_j \mathcal{N}_j^0 Z_j} \; \sinh\left(\frac{eU_0}{2k_B T}\right), \tag{39}$$

which reduces to

$$\sigma \approx \kappa\varepsilon\varepsilon_0\phi_0 \tag{40}$$

in the Debey–Hückel approximation, because $\sinh(x) \approx x$ for $x \ll 1$. Thus, the potential emerging from a flat wall with charge density σ immersed in a solution of micro–ions takes the simple form

$$U_W(D) = \frac{\sigma}{\varepsilon\varepsilon_0\kappa} \exp\{-\kappa D\}. \tag{41}$$

Note that this constitutes another asymptotic limit of eq. 36 for $R/D \gg 1$ and $\kappa R \gg 1$ with $D = r - R$.
For most practical purposes in soft matter physics the mean electrostatic potential around a single charged entity is of limited interest. The quantity of central interest is rather the pair interaction potential $\phi^{er}(r)$ between two charged objects. This is the potential energy of one charged object and its diffuse layer in the electric field of a second object with its diffuse layer, which here is

$$\frac{\phi_{PP}^{er}(r)}{k_B T} = L_B Z_P^2 \frac{1}{r} \exp\{-\kappa r\} \qquad \text{for two point charges at r} > 0 \tag{42}$$

$$\frac{\phi_{SS}^{er}(r)}{k_B T} = L_B Z_S^2 \left(\frac{\exp\{\kappa R\}}{1+\kappa R}\right)^2 \frac{1}{r} \exp\{-\kappa r\} \qquad \text{for two charged spheres with } \kappa R \lesssim 5 \text{ at } r > 2R \tag{43}$$

$$\frac{\phi_{WW}^{er,a}(D)}{k_B T} = \frac{L_B}{A^2\kappa} 4\pi Z_W^2 \exp\{-\kappa D\} \qquad \text{for two parallel walls of surface area } A \text{ at distance } D > 0. \tag{44}$$

Table 3: *Rules of thumb for the calculation of κ^{-1} for different types of electrolytes. The salt concentration, c_{salt} has to be introduced in units of mol/L*

Type of electrolyte	example	κ^{-1}/ nm
1:1	NaCl	$\frac{0.304}{\sqrt{c_{salt}}}$
1:2 or 2:1	$CaCl_2$	$\frac{0.176}{\sqrt{c_{salt}}}$
2:2	$MgSO_4$	$\frac{0.152}{\sqrt{c_{salt}}}$

Here the Bjerrum length $L_B = \beta e^2/4\pi\varepsilon\varepsilon_0$ is the distance at which the electrostatic energy between two elementary charges equals the thermal energy. For the case of two charged parallel walls $\phi_{WW}^{er,a}(D)$ is the electrostatic potential energy per unit area, because the total energy between parallel walls is proportional to their surface area. Further, this expression for the energy between two walls can be used to calculate the interaction potential profile between two spheres of radius R applying Derjaguin's approximation

$$\begin{aligned} \frac{\phi_{SS}^{er}(D)}{k_BT} &\approx 2\pi R\frac{\sigma^2}{\varepsilon\varepsilon_0 k_BT\kappa^2}\exp\{-\kappa D\} \\ &= \frac{L_BZ_S^2}{2R^3\kappa^2}\exp\{-\kappa D\} \end{aligned} \tag{45}$$

provided the thickness of the diffuse micro–ion layer is significantly smaller than the spheres' radius.

3.3 Electrolyte concentration

The screening length κ^{-1} is the key parameter for the quantitative description of electrostatic repulsion between colloidal particles, as influences the electrostatic energy in a complex way. It determines the the amplitude as well as the range of the potential. According to its definition, κ depends on the dielectric constant of the solvent and more importantly, it is proportional to the square root of the concentration and the valency of additional electrolytes present in the solution. If the electrolyte concentration, c_{salt}, is expressed in mol/L and the constants entering into κ are given in SI–units, one obtains the screening length in aqueous solutions from the rule of thumb

$$\kappa^{-1} = \frac{0.43}{\sqrt{\sum_j Z_j c_{salt}}} \quad [\text{nm}] \tag{46}$$

in units of nanometer. For some types of electrolytes the corresponding expressions are listed in Table 3. In an aqueous solution with a $NaCl$ concentration of 10^{-3} mol/l the screening length is approximately 10 nm.

4 The DLVO Theory and Colloidal Stability

When calculating the electrostatic potential between spherical particles, we have neglected van der Waals interactions completely, which is a gross simplification, since charges will of course

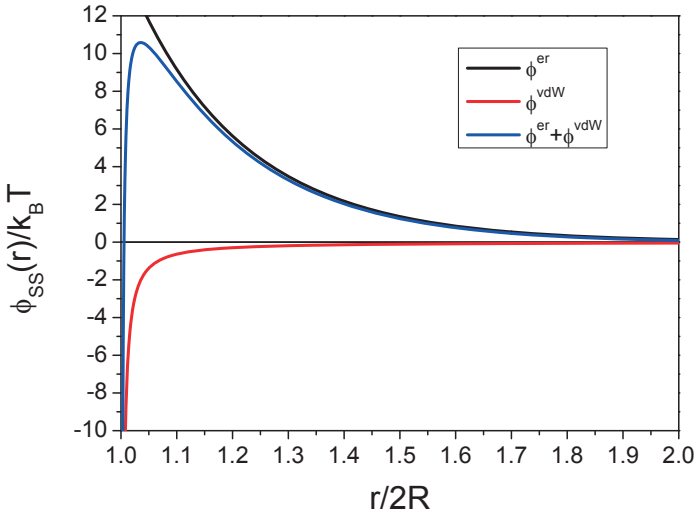


Fig. 5: *Contributions to the DLVO-interaction potential profiles between two spheres. Curves are calculated for two spheres with* $R = 250$*nm and* $A_H = k_BT$ *in water where* $L_B \approx 0.7$*nm. The screening length and the charge density are* $\kappa R = 2$ *and* $\sigma = 5 \times 10^{-3} e/R^2$

.

interact with permanent dipoles within the spheres. However instead of calculating the interactions of individual charges with the spheres, we assume that the ions are evenly distributed and consequently produce a net zero mean force onto the particles. Further the electrostatic and the van der Waals potential between the colloids are assumed to be additive. This is the basis of the so called DLVO–theory (named in the honor of Derjaguin, Landau, Vervey and Overbeeck) of colloidal stability. The basic features of DLVO-potentials and their implications for the stability of a colloidal suspension are illustrated in Figures 5 through 7, where we plot calculated potential profiles between two spheres with $R = 250$nm and $A_H = k_BT$ in water where $L_B \approx 0.7$nm.

Due to the fact that the electrostatic contribution increases exponentially on approaching the interface and the van der Waals potential decay with $1/D$, the latter will always dominate at very small distances. However, if the repulsion is strong enough the total interaction may have a high barrier, which prevents the particles from coming close enough for the van der Waals attraction to be effective. In this case the suspension is considered to be stable. The barrier may be lowered by either increasing the concentration of micro–ions by which the screening length is decreased (see Figure 6) or by decreasing the surface charge density σ (see Figure 7) of the particles by which the amplitude of the electrostatic repulsion term is reduced. In both cases the suspension is considered to become unstable if the barrier height drops to about k_BT.

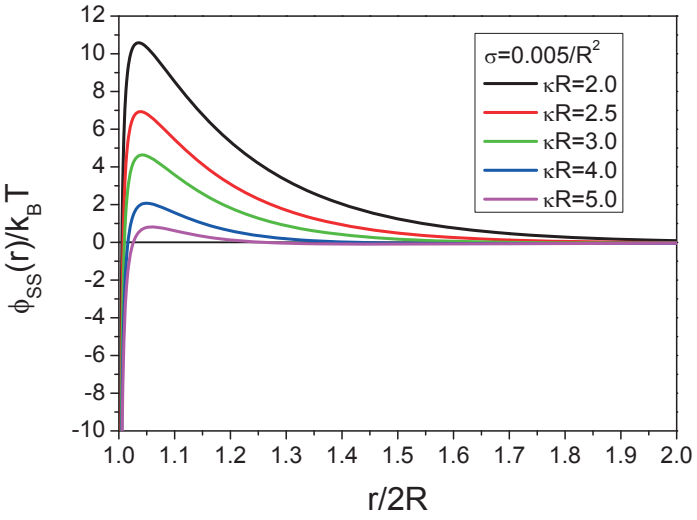


Fig. 6: *DLVO–potentials between two colloidal spheres at constant surface charge density with varying screening length as indicated in the legend. The particle parameters are the same as in Fig. 5*

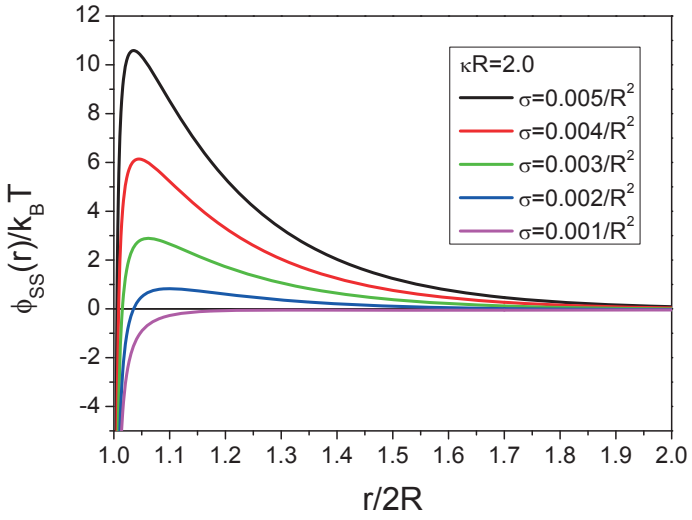


Fig. 7: *DLVO–potentials between two colloidal spheres at constant screening length with varying surface charge density as indicated in the legend. The other particle parameters are the same as in Fig. 5*

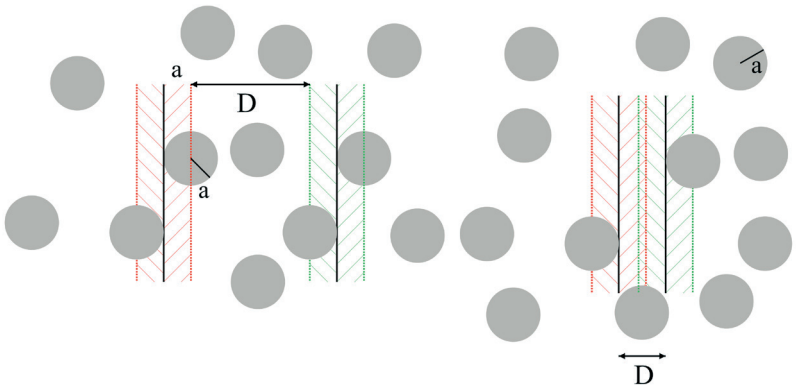


Fig. 8: *Two plates in a solution of spheres. At separation distances $D > 2a$ the small spheres can enter the gap between the plates (left). The spheres are depleted from the gap if $D \leq 2a$ (right)*

5 Non-DLVO interactions

5.1 Depletion interaction

If colloidal particles are dispersed in a solution which contains a further kind of solute, which has a typical length scale in the colloidal regime, an additional interaction has to be considered. Imagine two plates immersed in a solution containing small spheres of radius a as sketched in Figure 8. Then, the center of mass of any sphere can not get closer than a separation distance a to the plate. In other words in front of either plate there is a zone of thickness a which is depleted from spheres. The total volume, which for this is reason is not accessible to the spheres is called the excluded volume V_{ex}. We will see in the following that this depletion leads to an attractive interaction between the two plates, which for obvious reasons is referred to as depletion interaction. Here we will give a heuristic introduction to the subject. For an exhaustive treatment the reader is referred to the book by Tuinier and Lekkerkerker [13].
The effective interaction potential between the two plates W_{WW}^{depl} may be regarded as the change in Helmholtz free energy ΔH of the whole system, when the plates are brought from an infinite distance to their final separation D. For a system at constant temperature and composition, $\Delta H = -\Pi \Delta V$, with Π the osmotic pressure and ΔV the negative change of excluded volume, ΔV_{ex}. If we treat the spheres as an ideal gas, we may immediately write down an expression for the depletion interaction between the two plates.

$$\phi_{WW}^{depl} = \rho k_B T \left[V_{ex}(D) - V_{ex}(D \to \infty) \right] \tag{47}$$

where we we treat the depletant as phantom spheres, which do not interact by mutual excluded volume, such that $\Pi = \rho k_B T$ with ρ the sphere number density. The excluded volume at a given separation distance is easily calculated from geometrical considerations. If the plates are further apart than a sphere diameter

$$V_{ex}(D \to \infty) = 4Ar \tag{48}$$

where A is the area of a single plate surface (note that there are four surfaces). If the two plates get closer together then $D < 2a$ their depletion zones overlap and the spheres can not get into the gap between the plates. This leads to a density difference, which causes an osmotic pressure imbalance that presses the plates together. For the quantitative description of the resulting interaction potential we need to calculate $V_{ex}(D)$. Which is

$$V_{ex}(D) = 2Aa + DA \tag{49}$$

as can be easily seen from Fig.8. Note, that the gain in accessible volume, is equal to the volume where the two depletion zones overlap (see green and red hatched areas in Fig. 8, i. e. $V_{ol} = -\Delta V_{ex}$. The combination of eqs. 47, 48 and 49 gives

$$\frac{\phi_{WW}^{depl,a}(D)}{k_B T} = -\rho\,(2a - D)\,. \tag{50}$$

Although, this equation yields finite values for the interaction if $D > 2a$, it is important to note that this is not a physical meaningful result. From Fig. 8 it is evident that $\Delta V_{ex} = 0$, if $D > 2a$ and consequently eq. 47 results $W_{WW}^{depl} = 0$ in this case.
It is important to note that eq. 47 holds in general independently whether the surfaces under consideration are curved or not. Consequently there are two ways to calculate the depletion potential between two spheres. First we use the Derjaguin approximation for two spheres, which gives

$$\begin{aligned} \frac{\phi_{SS}^{depl}(D)}{k_B T} &\approx -2\pi \frac{R_1 R_2}{R_1 + R_2}\rho \int_D^{2a} dz\,(2a - z) \\ &= -2\pi \frac{R_1 R_2}{R_1 + R_2}\rho \left(2a^2 - 2aD + \frac{D^2}{2}\right). \end{aligned} \tag{51}$$

The upper limit for the integration has to be $2a$ because, the depletion interaction between two flat surfaces is zero beyond this separation distance.
It is very instructive to compare the approximate solution of eq. 51 to the exact solution, which one obtains, if the overlap volume between two spherical depletion zones is calculated from exact geometrical considerations. For two large spheres of equal radius R we get

$$V_{ol}^{SS} = -\Delta V_{ex}^{SS} = \frac{2\pi a^3}{3}\left(1 - \frac{D}{2a}\right)^2 \left(2 + \frac{3R}{a} + \frac{D}{2a}\right), \tag{52}$$

with which with we can immediately write

$$\begin{aligned} \frac{\phi_{SS}(D)}{k_B T} &= -\rho \frac{2\pi a^3}{3}\left(2 + \frac{3R}{a} + \frac{D}{2a}\right)\left(1 - \frac{D}{2a}\right)^2 \\ &= -2\pi\rho R \left(\frac{2a^3}{3R} + a^2 + \frac{Da^2}{6R}\right)\left(1 - \frac{D}{2a}\right)^2, \end{aligned} \tag{53}$$

For Large R/D and R/a this gives the same expression as eq. 51 for the case of $R_1 = R_2 = R$. The effect of the Derjaguin approximation is illustrated in Figure 9, where we compare depletion potentials between two spheres of equal radius R induced by a second species of spheres with radius $a < R$ at a density of small spheres $\rho = 0.1\rho^*$ with ρ^* the number density at close

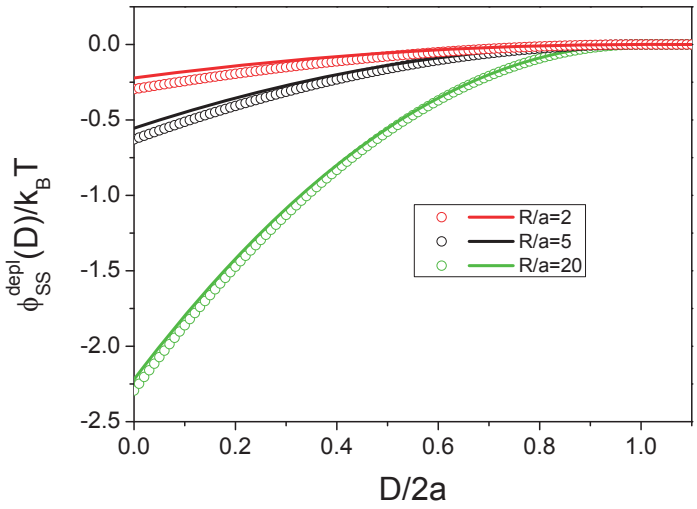


Fig. 9: *Calculated depletion potentials between two spheres of equal radius R induced by a second species of spheres with radius $a < R$, as indicated in the legend. Symbols represent exact calculations, while the full lines were calculated using Derjaguin's approximation. The number density of small spheres is $\rho = 0.1\rho^*$.*

sphere packing. If $R >> a$ the approximation gives good agreement with the exact solution, while for smaller spheres the approximation gives values which are significantly too small.
From eqs. 51 or 53 it can be seen that the depth of the potential depends on the number density ρ of the small spheres at constant R/a. This is shown in Figure 10. If the number density is fixed, the depth of the potential increases with R/a and the range of the potential increases with a in any case.

5.1.1 Non-spherical depletants

For non-spherical, particulate depletants, the orientational degrees of freedom are reduced upon the approach to the colloids' surfaces, which like in the case of polymeric depletants leads to a continuous variation of the depletant number density along the surface normal. In these cases the depletion potential is usually calculated in a two step procedure. First the potential or the force between two flat surfaces is derived again treating the depletant as an ideal gas, In a second step the potential between spherical particles is then calculated using Derjaguin's approximation. The depletion potential per unit area between two flat plates at distance D caused by thin rods of length L at low densities can be written as

$$\frac{\phi_{WW}^{depl,a}(D)}{k_BT} = \rho\left(D - \frac{D^2}{2L} - \frac{L}{2}\right) \tag{54}$$

For details of the derivation of this expression the reader is referred to the original work of Asakura and Oosawa[14, 15]. Applying Derjaguin approximation the interaction potential be-

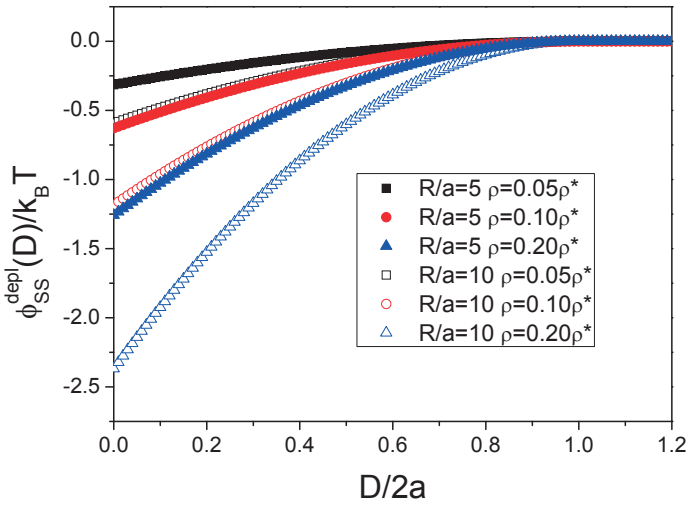


Fig. 10: *Depletion interaction between two spheres mediated by small spheres for different size ratios of the two species and different volume fractions of the small spheres*

tween two spheres of equal radius R is readily calculated to be

$$\frac{\phi_{SS}^{depl}(D)}{k_BT} = -\rho\pi R\frac{L^2}{6}\left(1-\frac{D}{L}\right)^3 \tag{55}$$

for separation distances smaller than the rod length. At $D \geq L$ the potential is zero.
The depletion interaction between two large spheres induced by oblate spheroids has first been investigated in the low density limit by Piech et al.[16]. Applying the Derjaguin approximation they find for the limiting case of infinitely thin circular discs of diameter d

$$\frac{\phi_{SS}^{depl}(D)}{k_BT} = \begin{cases} 0 & \text{for} D > d \\ -\rho\frac{\pi R}{3}d^2\left[\frac{3}{2}\frac{D}{d}\sin^{-1}\frac{D}{d} - \pi\frac{3}{4}\frac{D}{d} + \left(1+\frac{1}{2}\frac{D^2}{d^2}\right)\sqrt{1-\frac{D^2}{d^2}}\right] & \text{for} D \leq d \end{cases} \tag{56}$$

From eqs. 51, 55 and 56 it is evident, that the potential at $D = 0$, the so-called contact potential, $\phi_{SS}^{depl}(0)$ is proportional to $\rho R/l\rho^*$, where l is the is the maximum range of the depletion interaction, i. e. $2a$ for spheres, L for rods and d for discs while $\rho^* = 6/\pi l$ is the depletants overlap concentration. Quantitatively, we find that the contact potential caused by rod shaped depletants has the smallest absolute value, the absolute contact value due to discs is twice as large and that caused by spheres is three times larger, at the same ρ/ρ^*. This result, which is illustrated in the in the left part of Fig. 11, implies that spherical co-solutes are three times more efficient depletants than rods. However, for technical applications, the efficiency per mass of depletant is the relevant quantity. In the right panel of Fig. 11 we compare the depletion potentials calculated at constant mass concentration $c = \rho V_p \rho_m = 10$mg/mL where V_p is the volume of a single depletant particle, and ρ_m is its mass density. We used an aspect ratio of 100 for

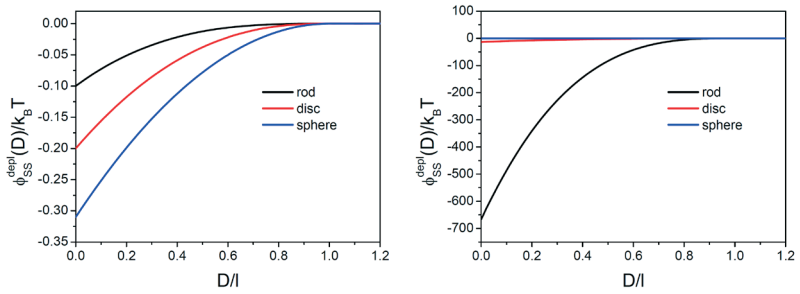


Fig. 11: *Comparison of depletion potentials between two spheres of radius $R = 1\mu m$ induced by differently shaped depletants, as indicated in the legends, with a ratio of $R/l = 10$. Left: potentials calculated at equal number density $\rho/\rho^* = 0.01$ of depletants. Right: potentials calculated at equal mass concentration of depletants.*

the rods and the discs and the same mass density $\rho_m = 1$g/mL for all three types of depletants. Now the rods are by almost two orders of magnitude more efficient than discs, which are another two orders of magnitude more effective than spheres. This is caused by the fact, that due to orientational degrees of freedom, the same amount of mass occupies much more volume in the case of anisotropic particles as compared to the case of spheres.

5.2 Steric repulsion

It has been known on a phenomenological level since very long time, that mixing suspensions of different materials can be used to stabilize liquid suspended matter. Ancient Egyptians manufactured ink slurries by mixing soot or iron oxide containing earth particles with aqueous arabic gum solutions. Without knowing, they exploited the fact that the polysaccharide adsorbs onto the particles' surfaces by that increasing their colloidal stability by steric repulsion. In a very coarse picture, this effect can be visualized as the anchored polymers by excluded volume interaction keeping the particles separated at distances where their van der Waals interactions are not any more strong enough to cause aggregation. There is no simple theory available which describes steric repulsion forces comprehensively. This is mainly due to the fact that a plethora of parameters determine the interaction potential, of which the three most prominent ones are the quality of the solvent for the polymer, the amount of polymer on the particle surface and the way in which it is attached. Here we will restrict the discussion to the situation where the polymers are chemically linked to the particle surface with a grafting density, which is $\Gamma >> 1/R_g$ with R_g the polymers' radius of gyration. This situation is usually referred to as the brush repulsion regime, and brush repulsion is another very effective tool to stabilize colloids against aggregation. Alexander [17] and de Gennes [18] have calculated the repulsive force per unit area between two planar surfaces, which is often termed the disjoining pressure, Π. If both surfaces are covered with a brush of hight H_b the dependence of the disjoining pressure on separation distance is

$$\Pi^{sr}(D) = k_B T \Gamma^{\frac{3}{2}} \left[\left(\frac{2H_b}{D} \right)^{\frac{9}{4}} - \left(\frac{D}{2H_b} \right)^{\frac{3}{4}} \right] \tag{57}$$

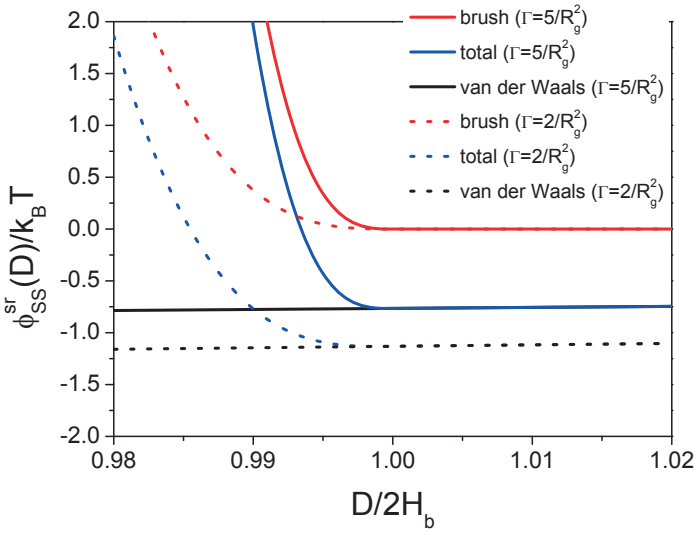


Fig. 12: *Interaction profiles between two spheres with radius* $R = 100nm$ *and* $A_H = 1k_BT$ *covered with a brush consisting of polymers with* $n_s = 10$ *segments of length* $l_s = 0.25nm$ *calculated for two different grafting densities. Full lines correspond to a grafting density* $\Gamma = 5/R_g^2$ *while dotted lines represent* $\Gamma = 2/R_g^2$.

In this situation, the height of the brush is approximately given by $H_b \approx n_s l_s^{5/3} \Gamma^{1/3}$ where n_s is the number of segments in the polymer chains and l_s is their length. Using Derjaguin approximation the corresponding interaction potential is obtained as [19, 20]

$$\frac{\phi_{SS}^{sr}(D)}{k_BT} = \begin{cases} 0 & \text{for} D > H_b \\ \frac{16\pi a H_b^2 \Gamma^{\frac{3}{2}}}{35}\left[28\left(\left(\frac{2H_b}{D}\right)^{\frac{1}{4}} - 1\right) + \frac{20}{11}\left(1 - \left(\frac{D}{2H_b}\right)^{\frac{11}{4}}\right) + 12\left(\frac{D}{2H_b} - 1\right)\right] & \text{for} D \leq H_b \end{cases} \tag{58}$$

From Fig.12 it is evident that brush repulsion can be a very effective stabilization mechanism. There we show the pair interaction potential between two spheres with radius $R = 100$nm and Hamaker constant $A_H = 1k_BT$ covered with a brush consisting of short oligomers with $n_s = 10$ segments of length $l_s = 0.25$nm corresponding to a radius of gyration $R_g \approx 0.3$nm. At a grafting density of $\Gamma = 5/R_g^2$ the a brush height is $H_b \approx 3.6nm$ and the steric repulsive rises extremely steeply. If the particles were forced to a separation distance of $D = 1.9H_b$ the steric repulsion raise to $\phi_{SS}^{sr} \approx 350k_BT$. Even for a lower grafting density of $\Gamma = 2/R_g^2$ where the brush height reduces to $H_b \approx 2.7$nm the repulsive potential is $\phi_{SS}^{sr} \approx 50k_BT$ at a separation distance of $D = 1.9H_b$.

6 Experimental Methods

As is discussed in the previous sections the interaction energies between a pair of colloidal spheres is of the order of several thermal units. Therefore, experimental investigations of these

pair potentials require extremely sensitive methods. To date the only methods which provide the necessary force resolution are optical tweezers experiments in combination with high resolution video microscopy [21] and total internal reflection microscopy (TIRM) [22]. Both methods rely on the same basic idea, that the histogram of separation distances between two bodies in thermal equilibrium is a good approximation for the probability density of separation distances $p(D)$, from which the potential energy between the two bodies is calculated using Boltzmann's law. In situations where the a very good spatial resolution is required rather than a force sensitivity, the methods of choice are either the so called surface forces apparatus (SFA) [23] or chemical probe atomic force microscopy [25]. In the following we will discuss SFA and TIRM in some detail.

6.1 Surface forces apparatus

The Surface Forces Apparatus, is an instrument, which allows to measure forces in the range of some pico-Newton occurring between two macroscopic surfaces. It was first developed by Israelachvili[23] in the late seventies for the purpose of measuring van der Waals and electrostatic interactions. Over the years it was further developed for the investigation of all kind of colloidal interactions, which were discussed in the previous sections. An extensive list of studies on depletion and steric repulsion forced induced by polymers, was given by Kleshchanok et al [24].

A schematic sketch of an SFA is shown in Fig.13. The core part consists of two hemi-cylindrical lenses with a curvature radius of the order of 1 cm, which are mount with their cylinder axis intersecting at an angle of ninety degree. One of the two cylinders is supported by a system of springs with different stiffness, while the other can be very precisely positioned by a piezo actuator. The distance, D, between the two curved surfaces can be changed by a value M, moving the actuator. When the surfaces are separated by a distance which is larger than the range of any acting forces, a motion of the actuator will change the surface separation by the same amount, i. e. $M = \Delta D$. Since the other cylinder is mounted on a compliant spring, the situation is different when a surface force, $F(D)$, deflects the spring. In this case the equilibrium distance will adjust to balance the force acting between the surfaces and the restoring spring force, where the spring with the lowest stiffness dictates the deflection. From the deviation of ΔD from M, the force acting between the two surfaces can be directly obtained from Hook's law $F(D) = k_1(M - \Delta D)$. Thus, it is of utmost importance to measure the surface separation distance accurately, which is achieved by a technique called Multiple Beam Interferometry (MBI). White light is directed at normal incidence through the crossed cylinders, which are typically made of quartz to which mica sheets are glued. The latter have a thickness of 2-5 μm and covered with a silver coating with a thickness of several tenths of nanometers on their backsides. The silver coatings act as semi-transparent mirrors between which the white light is reflected multiple times before leaving the gap between the cylinders. When the separation distance between the silver coatings is an integer multi-fold of a particular wavelength, constructive interference is observed. Therefore, in a spectrometer attached to the SFA, fringes of high intensity are observed for these particular wavelengths. From the positions of the fringes of equal chromatic order it is possible to determine the separation distance D with a resolution of ≈ 0.1 nm.

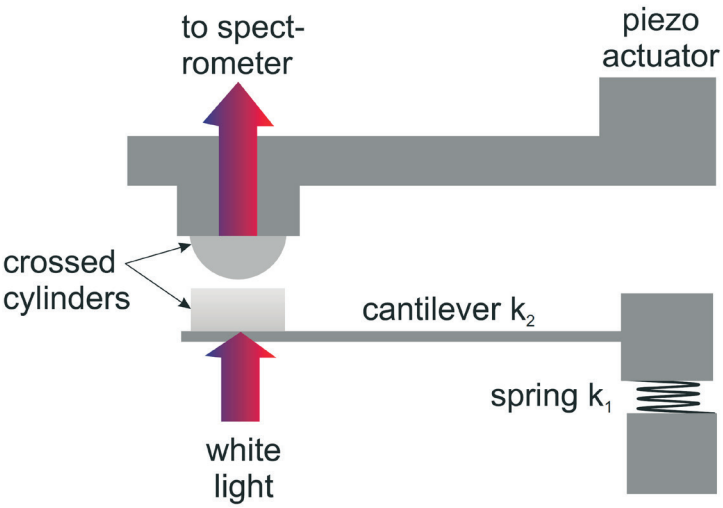


Fig. 13: *Sketch of working principle of a surface forces apparatus. The entire assembly can be immersed in a liquid suspension for measurement of colloidal forces.*

6.2 Total internal reflection microscopy

As seen in the previous sections the SFA provides an excellent distance resolution and the smallest forces which can be measured are in the pico-Netwon range. In this sense, total internal reflection microscopy (TIRM) is a complementary method. It applies an optical gauge to measure interaction potentials, which allows to determine forces down to the femto-Newton range. On the other hand the distance resolution in TIRM is limited to roughly 1 nm. The basic principle of TIRM relies on the condition that a colloidal sphere has to undergo Brownian motion in a shallow potential well close to a flat interface. This situation can be observed, e. g. if a charged colloidal sphere in solution, which is large enough to sediment in the field of earth gravity eventually comes close enough to the like charged bottom wall of the sample cell. Then, the superposition of gravitational contribution and electrostatic repulsion leads to an interaction potential between the particle and the wall with a shallow minimum at a distance D_{min}. Due to Brownian motion the particle will sample a distribution of separation distances around the minimum, $p(D)$, which is related to the interaction potential profile by Boltzman's equation

$$p(D) = A \exp\left\{-\frac{\phi(D)}{k_B T}\right\}. \tag{59}$$

Here $p(D)dD$ is the probability to find the particle in the height interval $D + dD$ and A is a constant normalizing the integrated distribution to unity.

The fluctuations of separation distance resulting from the thermal motion can be directly observed by TIRM. For this purpose a Laser beam is directed to the container/solution interface via a prism to enable an angle of incidence, α_i, at which total reflection occurs. According to Maxwell's equations the electric field of the Laser beam can not change discontinuously at

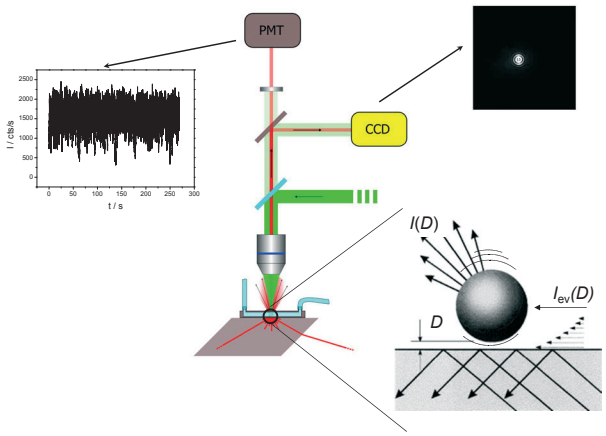


Fig. 14: *Sketch of a TIRM setup. By total reflection of the red laser from the glass/liquid interface the evanescent wave is created. A colloidal sphere situated in the evanescent wave scatters light with an intensity which depends on the the sphere to interface separation distance. The scattered signal can be detected either with a photo multiplier tube or a CCD-camera. The particle is prevented from moving out of the field of observation by weak optical tweezers, which is created by coupling a differently colored laser into the back aperture of the microscope objective.*

the reflecting interface. It rather penetrates the interface causing an evanescent wave with an amplitude which decays exponentially with the distance from the interface. The distance where the field strength has decayed to $1/e$ of the original strength is termed penetration depth Λ. This is given by $\Lambda = 2\pi/\lambda_0\sqrt{(n_1 \sin\alpha_i)^2 - n_2^2}$ with λ_0 the laser vacuum wavelength, n_1 and n_2 the index of refraction of the glass and the solvent, respectively.
A single colloidal sphere, which is situated in this evanescent wave will scatter light with an intensity $I_S(D)$ depending on its distance from the interface [22] as

$$I_S(h) = I(D=0)exp\{-\Lambda D\}, \tag{60}$$

where Λ is the inverse penetration depth of the evanescent wave. Therefore, the particle's separation distance fluctuations will cause fluctuations of the scattered intensity and

$$p(I_S(h))dI_S(h) = p(h)dh \tag{61}$$

where $p(I_S(D))dI_S(D)$ is the probability to observe scattered intensity in the interval $I_S(D) + dI_S(D)$. Introducing eqs. 60 and 59 into eq. 61 relates the the probability density of the scattered intensity to the interaction potential by

$$-p(I_S(D))\frac{I_S(D=0)}{\Lambda}exp\{-\Lambda D\} = A\exp\left\{-\frac{\phi(D)}{k_BT}\right\}. \tag{62}$$

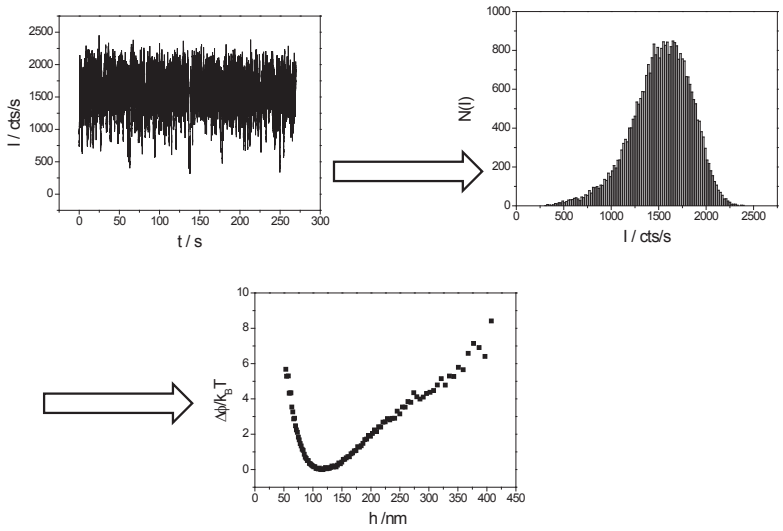


Fig. 15: *Sketch of working principle of total internal reflection microscopy. The trace of scattered intensity over time is converted to a histogram. In the case of a sufficiently large number of events, this is a good approximation for the probability distribution to find a particle at a given separation distance from the interface. From this the potential profile is calculated using Boltzmann's law*

A TIRM set up is sketched in Fig. 14. In a typical experiment scattered intensities are recorded with a time resolution in the range of 2-10 ms, which corresponds to an intensity trace over time with about $2 \cdot 10^5$ data points recorded in fifteen minutes. Therefore the normalized intensity histogram $N(I_S(D))$ may be regarded as a good approximation for the probability distribution $p(I_S(h))$. Normalizing eq. 62 by the corresponding expression for the intensity which is observed with the highest frequency, $I_S(D_m)$ yields

$$\frac{N(I_S(D))}{N(I_S(D_m))}\frac{I_S(D)}{I_S(D_m)} = \exp\left\{-\frac{\phi(D)-\phi(D_m)}{k_BT}\right\}. \tag{63}$$

By solving eq. 63, the potential difference $\Delta\phi = \phi(D) - \phi(D_m)$ can be expressed as a function of four measurable quantities

$$\frac{\Delta\phi}{k_BT} = \ln\left(\frac{N(I_S(D_m))}{N(I_S(D))}\right) + \ln\left(\frac{I_S(D_m)}{I_S(D)}\right). \tag{64}$$

The second logarithmic term on the right hand side of eq. 64 is related to the separation distance by eq. 60, which yields

$$\frac{\Delta\phi}{k_BT} = \ln\left(\frac{N(I_S(D_m))}{N(I_S(D))}\right) + \Delta D\Lambda \tag{65}$$

where $\Delta D = D - D_m$ and Eq. 65 represents the working instruction how to measure the potential profile between a colloidal spheres and a flat solid interface by TIRM, which is summarized

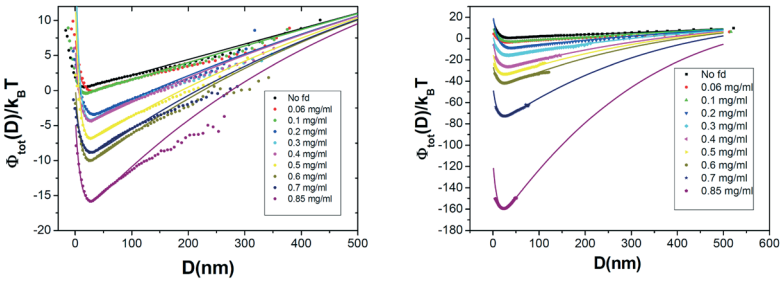


Fig. 16: *Interaction potentials between a probe sphere with $R = 0.5\mu m$ (left) and $R = 2.0\mu m$ (right) measured at different concentrations of fd-virus in the suspending solution.*

in Fig. 15 . In order to obtain potential profiles on an absolute scale of separation distances,the scattered intensity $I(D = 0)$ from a particle which was forced to sediment onto the cell wall has to be measured. With the knowledge of $I(D = 0)$ scattered intensities can be translated to absolute distances by resolving eq. 60 for D.

6.2.1 Depletion interactions measured with TIRM

As mentioned in section 5.1 two basic assumptions were applied to derive a closed analytical expressions for the depletion potential between colloidal entities induced by rod shaped particles. The depletant is treated as an ideal gas and Derjaguin approximation is assumed valid. In this section we will discuss a TIRM study where we investigated the limits of these approximations [26]. For this purpose interaction potentials between probe spheres with different radii R and a glass wall were measured. Depletion interaction was induced by the addition fd-virus, which is a bacteriophage with a length $L_{fd} = 880$nm and a cross section radius of 7 nm. By increasing the size ratio between probe sphere and depletant up to $L_{fd}/R = 1.76$ and the virus concentration to $\rho \approx 10 \times \rho^*$ the two approximations were violated willingly.
Experimental data for the smallest and the largest sphere radius are displayed in Fig. 16 together with the best fits to a model function representing the superposition of eqs. 45 and 55 with a gravitational contribution given by

$$\frac{\phi_{tot}(D)}{k_B T} = B \exp\{-\kappa D\} - \frac{c_{fd} N_A \pi}{3 M_{fd}} {L_{fd}}^2 R \left(1 - \frac{D}{L_{fd}}\right)^3 - \frac{F_{app} D}{k_B T} \tag{66}$$

. Here, all the parameters determining the strength of the electrostatic repulsion are condensed into the amplitude B, the rod concentration is given in units of mass per volume $c_{fd} = \rho_{fd} M_{fd}/V$ with the virus' molar mass, M_{fd} and Avogadro's number, N_A. Finally the the force F_{app} is the some of the probe sphere's buoyancy corrected weight force and the photon pressure of the optical tweezers. In this model function, all the parameters are known or can be calibrated by independent experiments. Thus any deviation of the best fitting parameter values from the a priory known values indicates a failure of the applied model. In Fig. 17 the fitted valued for the rod concentration c_{FIT} are plotted versus the actual rod concentration c_{SAMP}, which was determined by UV/VIS-spectroscopy. In this representation a linear relation with zero intercept and unity slope was expected, if the model described the experiment correctly throughout the entire range of concentrations.

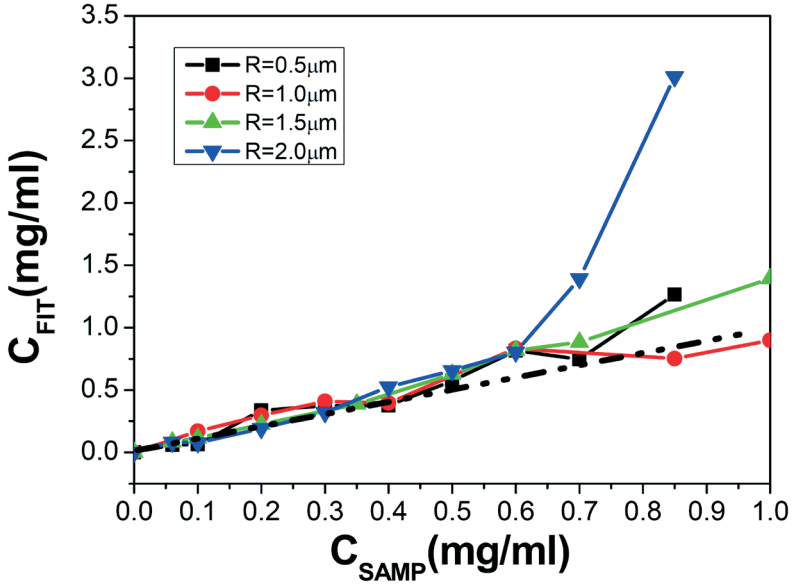


Fig. 17: *Concentrations values obtained from model fitting versus the samples' rod concentration as determined by UV/VIS spectroscopy. Different symbols represent different probe sphere sizes as indicated in the legend. The dashed dotted line represents a linear relation with zero intersect and unity slope.*

We observed that the experimental data follow the expected trend almost perfectly, independently of the probe spheres' diameter up to $c_{fd} = 0.5$mg/mL, which corresponds to $\rho/\rho^* \approx 7$. The data obtained with the two smallest probe spheres appear to follow the linear trend up to the highest concentration $\rho/\rho^* \approx 15$ within experimental scatter. Only for the case of the $R = 2\mu$m sphere and to a smaller extend for the there $R = 1.5\mu$m sphere significant deviation from the model predictions are observed at $c_{fd} > 0.5$ mg/mL. Here the experimental data deviate systematically from the predictions towards higher values which indicates a stronger attractive interaction than expected from the depletion model which applies Derjaguin and low density approximation. These observations allow an assessment of the limits of the two basic approximations. In those cases where Derjaguin approximation is expected to hold, i.e. for the spheres with $R = 1.5\mu$m and $R = 2\mu$m, the ideal gas approximation appears to break down only at fd-number densities as high as seven times the over-lap value. At higher depletant content, the attractive potential is significantly deeper than predicted. Numerical calculations reveal that deviations of experimental depletion potentials from the Derjaguin description are of the order of ten percent for $L_{fd}/R = 0.88$ [27, 28, 29] while the deviation increases to a factor of two for $L_{fd}/R = 1.76$. Differently, here in both cases no significant discrepancy was observed between the experimental data and the approximate theoretical prediction. It is however possible that this is due to a fortuitous balancing of two effects. While a violation of the ideal gas model is expected [31] to cause a deepening of the depletion potential, as was observed here with the large probe spheres, the violation of Derjaguin approximation is expected to have the opposite effect.

References

[1] E. S. Hedges *Colloids* (Edward Arnold & Co. London 1931)

[2] R. G. Jones, E. S. Wilks, W. V. Metanomski, J. Kahovec, M. Hess, R. Stepto, T. Kitayama, ed. *Compendium of Polymer Terminology and Nomenclature, IUPAC Recommendations 2008* (RSC Publ. 2009)

[3] B. Derjaguin, L. Landau *Acta Physico Chemica URSS*, **14** 633 (1942).

[4] E. J. W. Verwey, J. T. G. Overbeek, *Theory of the stability of lyophobic colloids* (Elsevier, Amsterdam, 1948).

[5] B. V. Derjaguin *Kolloid Z.*, **69** 155 (1934).

[6] for a comprehenive introduction see J. N. Israelachvili, *Intermolecular and Surface Forces 2nd ed.* (Academic Press, London, 1991).

[7] R. Eisenschitz, F. London, 491 1930

[8] F. London*Z. Physik*, **63** 245 (1930).

[9] F. London *Z. Physik. Chemie*, **B11** 222 (1930)..

[10] E. M. Lifshitz *Soviet Phys. JETP*, **2** 73 (1956).

[11] R. J. Hunter, *Foundations of Colloidal Science, 2nd ed.* (Oxford University Press, Oxford, 2001).

[12] J. Lyklema, H. P. van Leeuwen, M. Vliet, A.-M. Cazabat (eds.), *Fundamentals of Interface and Colloid Science: Particulate colloids*, (Elsevier, London, 2005).

[13] R. Tuinier, H. N. W. Lekkerkerker, *Colloids and the depletion interaction*. Springer, Lecture note in Physics 833, 2011

[14] S. Asakura, F. Oosawa *J. Chem. Phys.*, **22** 1255 (1954).

[15] S. Asakura, F. Oosawa *J. Polym. Sci.*, **33** 183 (1958).

[16] M. Piech, J. Y. Walz *J. Col. Interf. Sci.*, **232** 86 (2000).

[17] S. Alexander *J. Phys. (Les Ulis, Fr.)*, **38** 983 (1977).

[18] P. G. de Gennes *C. R. Acad. Sci. Ser. II, Fascicule B Mechanique Physique Chimie Astronomie*, **300** 839 (1985).

[19] C. N. Likos, K. A. Vaynberg, H. Löwen, N. J. Wagner *Langmuir*, **16** 4100 (2000).

[20] D. Kleshchanok, P. R. Lang *Langmuir*, **23** 4332 (2007).

[21] J. C. Crocker, D. G. Grier *Journal of Colloid and Interface Science*, **179** 298 (1996).

[22] D. C. Prieve *Colloid Interface Sci.*, **82** 93 (1999).

[23] J. N. Israelachvili, G. E. Adams *J. Chem. Soc. Faraday Trans. I*, **74** 975 (1978).

[24] D. Kleshchanok, R. Tuinier, P. R. Lang *J. Phys., Condens. Matter*, **20** 073101 (2008)..

[25] W. A. Ducker, T. J. Senden, R. M. Pashley *Nature*, **353** 239 (1991).

[26] S. De Sio, P. R. Lang *Z. Phys. Chem*, accepted for publication (2015).

[27] K. Yaman, M. Jeng, P. Pincus, C. Jeppesen, C. M. Marques *Physica A*, **247** 159 (1997).

[28] K. Yaman, C. Jeppesen, C. M. Marques, C. M. *Europhysics Letters*, **42** 221 (1998).

[29] P. R. Lang *J. Chem. Phys.*, **127** 124906 (2007).

[30] L. Auvray *Journal de Physique*, **42** 79 (1981).

[31] Y. Mao, M. E. Cates, H. Lekkerkerker *J. Chem. Phys.*, **106** 3721 (1997).

B 5 Formation of Polymer membranes - Turning polymer solutions into functional porosity

M. Wessling

DWI Leibniz Institute für Interaktive Materialien

Forckenbeckstrasse 50, 52062 Aachen

Contents

1 Introduction **2**
- 1.1 Applications . . . 2
- 1.2 The filtration spectrum . . . 2
- 1.3 Membrane morphologies . . . 3

2 Polymeric membrane formation **4**
- 2.1 Thermally induced phase separation . . . 5
- 2.2 Diffusion induced demixing . . . 5
- 2.3 Kinetics during membrane formation . . . 6
- 2.4 Integrally skinned membranes . . . 8
- 2.5 Simulations of membrane morphology . . . 8

3 Membrane fabrication processes **9**
- 3.1 Flat membranes . . . 9
- 3.2 Hollow fibers . . . 10

4 Summary **10**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

1.1 Applications

Technical membranes have emerged over the past 30 years in various areas; often unrecognized, they are either consumer products or essential tools in industries such as chemical industry, bioprocessing, gas separation and water production.
A few example illustrate their importance.

- Membranes are frequently used in clothing and shoes. The growing market of outdoor clothing relies on the function of synthetic porous as well as non-porous membranes to transport water vapor (sweat) from the atmosphere between human body and clothing to the outside environment. Such membranes prevent liquid water (rain) to penetrate the membrane. item Membranes are used in food protection. Food wrap has the function to encapsulate food to prevent moisture loss from the food matrix as well as to prevent the food to be continuously exposed to an oxygen rich atmosphere. Frequently a three-layer membrane is used having an oxygen barrier material laminated between two hydrophobic materials with low water permeability. The oxygen barrier must be kept water-free by the outside layers to prevent accelerated oxygen transport due to water swelling. item Reverse osmosis membranes have life support functions in arid regions. Countries in the Middle East produce about 50% of their drinking water by reverse osmosis. Membranes desalinate seawater by permeating water but rejecting the salt with retentions higher than 99%.

- Ultrafiltration and RO membranes produce drinking water. Currently, the drinking and process water production processes undergo a dramatic change: classical processing routes are replaced by membrane processes utilizing ultrafiltration as a pretreatment process and removing salts and small organic molecules by reverse osmosis.

- Membranes are major processing aids in beer and wine production. Clarification of wine by membranes is state-of-the-art. Dialysis membranes and reverse osmosis are used to produce alcohol-free beer. Recently, membrane separation processes penetrated the most conservative part of the brewing process: the water supply process.

- Membranes are major processing aids in dairy production. Microfiltration and ultrafiltration units belong to the standard unit operations to concentrate liquid streams present during milk and cheese processing.

1.2 The filtration spectrum

Figure 1 summarizes the application range of the various technical membrane filtration processes. These membranes are for more than 95% use of polymeric materials. Most of the functionality stem from their intricate porosity with pore size ranging from microns to subnanometer. This lecture will focus on the preparation of such membranes. It addresses the general principles and shows typical morphologies developing during the formation processes.

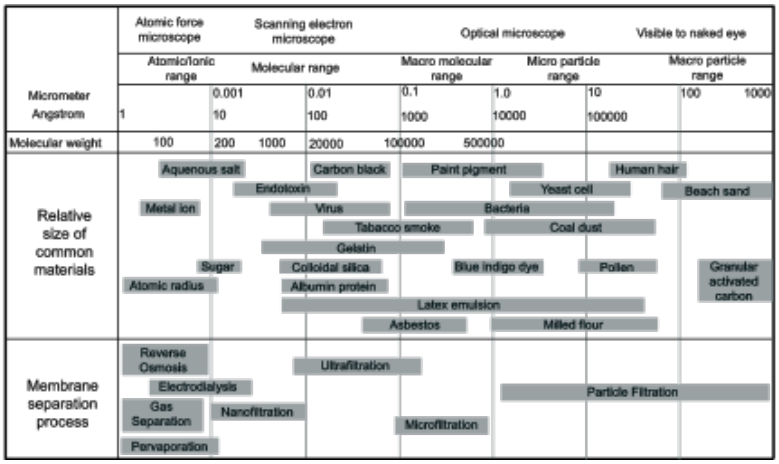


Fig. 1: *A size representation and the application range of various membrane processes*

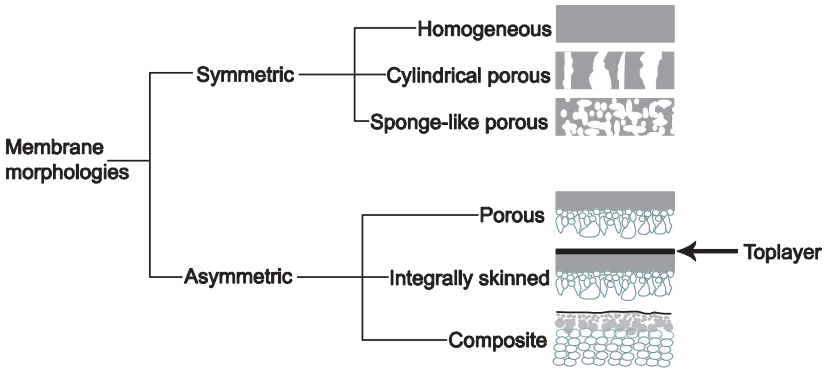


Fig. 2: *Representation of different membrane morphologies*

1.3 Membrane morphologies

One way of classifying membranes is by morphology or structure as shown in Figure 2. This is very illustrative because the membrane structure determines the separation mechanism and hence the application. If we restrict ourselves to solid synthetic membranes, two types of membrane may be distinguished, i.e. symmetric or asymmetric membranes. The two classes can be subdivided further as shown schematically in Fig . The thicknesses of symmetric membranes (porous or nonporous) range roughly from 10 to 300 μm, the resistance to mass transfer being determined by the total membrane thickness. A decrease in membrane thickness results in an increased permeation rate.

A breakthrough to industrial applications was the development of asymmetric membranes. These consist of a very dense toplayer or skin with a thickness of 0.1 to 0.5 μm supported by a porous sublayer with a thickness of about 50 to 250 μm. These membranes combine the high selectivity of a dense membrane with the high permeation rate of a very thin membrane. Figure 3 depicts the cross-section of a so-called multi-bore capillary hollow fiber with a very rich morphology in which the structural asymmetry is clearly visible. Large voids, larger pores

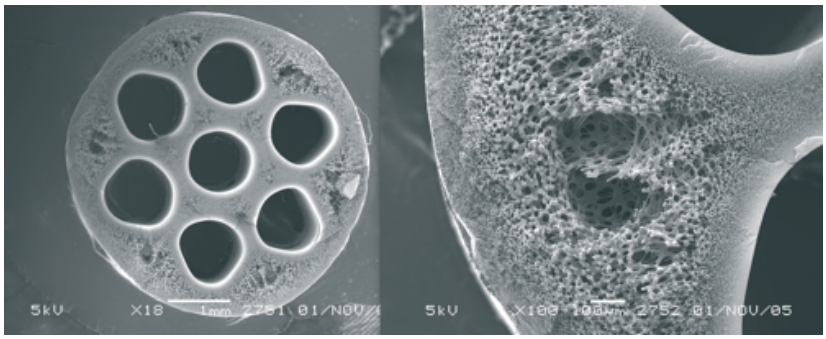


Fig. 3: *Cross-section of a multibore capillary membrane showing 7 lumen channels (left) and the complex morphology with pore sizes ranging from microns to nanometers (right)*

and decreasing pore sizes are visible. The resistance to mass transfer is determined largely or completely by a thin skin layer on the inside of the lumen: the skin is not visible at this small magnification. Its pore size in the order of 10 nm.
Many of the aspects summarized here are comprehensively described in Mulder's Book Basic Principles of Membrane Technology [1]

2 Polymeric membrane formation

Polymeric membranes are mostly formed from a polymer solution. The polymer solution is shaped into the desired configuration followed by a fixation (phase inversion) process, which determines the ultimate membrane structure and therefore the separation characteristics of the membrane.
The concept of phase inversion or phase separation covers a range of different techniques from solvent evaporation, precipitation by controlled evaporation, thermal precipitation, precipitation from the vapour phase to immersion precipitation. The majority of the phase inversion membranes are prepared by immersion precipitation which will be described in more detail. A comprehensive review is given in [2]
This paragraph mainly focuses on the phase inversion process. At first a more theoretical explanation of the phase inversion process is given. This part is split-up in two, one dealing with phase inversion induced by a change in temperature (Temperature Induced Phase Separation) and the other dealing with phase inversion induced by a change in composition (Diffusion Induced Phase Separation). The sections focus on the thermodynamics of binary and ternary polymer systems.
Phase inversion is a process that occurs far from equilibrium: a polymer is transformed in a controlled manner from a liquid to a condensed/solid state. The process of solidification is very often initiated by the transition from one liquid state into two liquids (liquid-liquid demixing). At a certain stage during demixing, one of the liquid phases (the polymer rich phase) will solidify so that a solid matrix is formed. By controlling the initial stage of phase transition the membrane morphology can be controlled, i.e. porous as well as nonporous membranes can be prepared. All phase inversion processes are based on the same thermodynamic principles, since the starting point in all cases is a thermodynamically stable homogeneous solution, which is subjected to a thermodynamic disturbance inducing demixing. Whether a mixture is thermody-

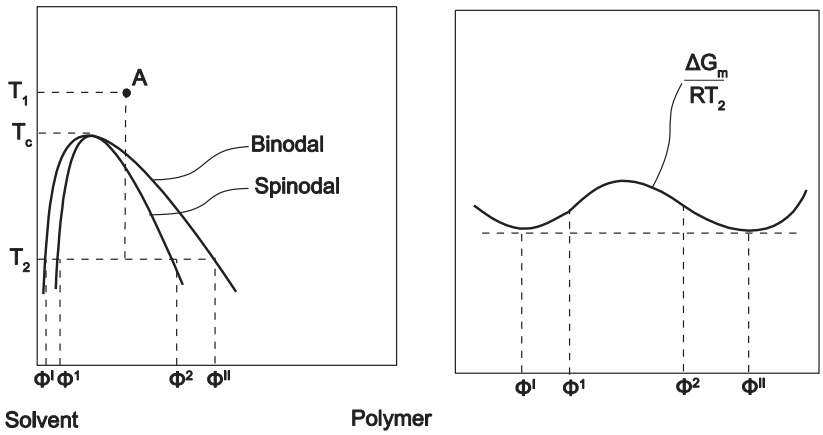


Fig. 4: *Phase diagram of a polymer solvent mixture showing the liquid liquid demixing region based on the function of the Free Enthalpy of Mixing* ΔG_m

namically stable or likes to demix into two separate phases is described by the free enthalpy of mixing, ΔG_m as shown in Figure 4 for a binary solution subjected to a disturbance of temperature quench.

2.1 Thermally induced phase separation

In order to understand the mechanism of liquid-liquid demixing more easily, a binary system consisting of a polymer and a solvent will be considered. Generally, the solvent has a high boiling point, e.g. sulfolane (tetramethylene sulfone, bp: 287 C) or oil (e.g. nujol). The starting point for preparing phase inversion membranes is a thermodynamically stable solution, for example one with the composition A at a temperature T_1 (with $T_1 > T_c$). All compositions with a temperature $T > T_c$ are thermodynamically stable in Figure 10. When the binodal is attained by decreasing the temperature, liquid-liquid demixing occurs and the solution separates into two phases, one rich in polymer and the other poor in polymer. When the temperature is decreased further to $T2$, the composition of the two phases follow the binodal and eventually the compositions ϕ_I and ϕ_{II} are obtained. At a certain temperature the polymer-rich phase solidifies by crystallisation (polyethylene), gelation (cellulose acetate) or on passing the glass transition temperature (atactic polymethylmethacrylate). Frequently, semi-crystalline polymers are used (polyethylene, polypropylene, aliphatic polyamides) which crystallise relatively fast, and hence a solid-liquid phase transition should be included. In case of glassy amorphous polymers the melting line may be replaced by a vitrification line.

2.2 Diffusion induced demixing

In addition to temperature changes, changes in composition brought about by the addition of a third component, a nonsolvent, can also cause demixing. Under these circumstances we have a ternary system consisting of a polymer, a solvent and a nonsolvent, where the solvent and nonsolvent must be miscible with each other. The liquid-liquid demixing area must now be represented as a three-dimensional surface. The free enthalpy of mixing is a function of the

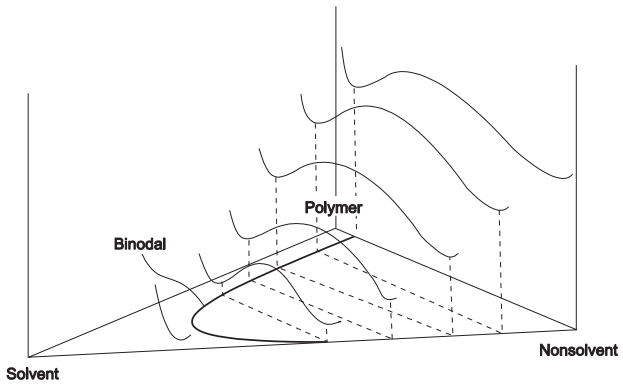


Fig. 5: *Phase diagram and ΔG_m representation for a ternary mixture of polymer/solvent/nonsolvent showing the liquid liquid demixing region based on the function of the free enthalpy of mixing*

composition as can be seen from Figure 5, where some drawings from the ΔG_m surface has been given at a certain temperature. All pairs of compositions with a common tangent plane to the ΔG_m surface constitute the solid line projected in the phase diagram, the binodal. The tie lines connect points on the binodal that are in equilibrium. A composition within this two-phase region always lies on a tie line and splits into two phases represented by the two intersections between the tieline and the binodal. As in the binary system, one end point of the tieline is rich in polymer and the other end point is poor in polymer. The binodal may be calculated numerically using various thermodynamic models. Most commonly, a Flory-Huggins type of model is used.

2.3 Kinetics during membrane formation

Most commercially available membranes are prepared by immersion precipitation. Immersion precipitation membranes in their most simple form are prepared in the following way. A polymer solution consisting of a polymer (3) and a solvent (2) is cast as a thin film upon a support (e.g. a glass plate) and then immersed in a nonsolvent (1) bath. The solvent diffuses into the coagulation bath (J_2) whereas the nonsolvent will diffuse into the cast film (J_1). After a given period of time the exchange of solvent and nonsolvent has proceeded so far that the solution becomes thermodynamically unstable and demixing takes place. Finally a solid polymeric film is obtained with an asymmetric structure. A schematic representation of the film/bath interface during immersion is shown in Figure 6. Precipitation occurs because of the exchange of solvent and nonsolvent. The membrane structure ultimately obtained results from a combination of mass transfer and phase separation.

Important questions arise for this non-equilibrium process: what factors are important in order to obtain a desired (asymmetric) morphology after immersion of a polymer/solvent mixture in a nonsolvent coagulation bath? Other interesting questions are: why a more open (porous) top layer is obtained in some cases whereas in other cases a very dense (nonporous) top layer supported by an (open) sponge-like structure develops? To answer these questions and to promote an understanding of the basic principles leading to membrane formation via immersion precipitation, qualitative description have been developed in the past. For the sake of simplicity, the

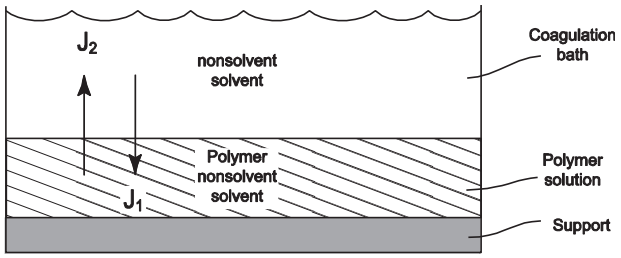


Fig. 6: *Schematic representation of a film/bath interface. Components: nonsolvent (1), solvent (2) and polymer (3).* J_1 *is the nonsolvent flux and* J_2 *the solvent flux*

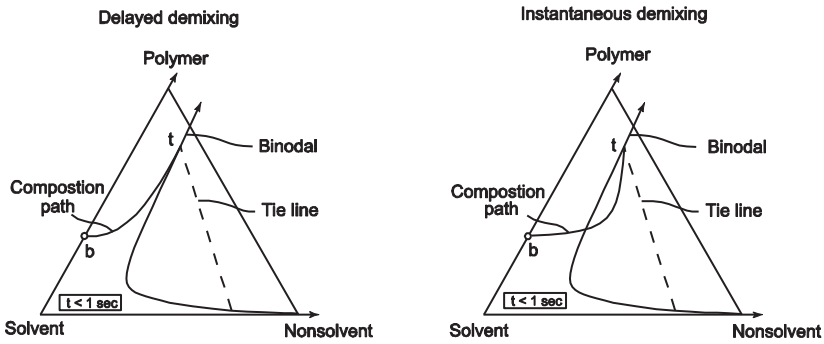


Fig. 7: *Representation of composition paths over the film thickness. The right-hand figure shows instantaneous liquid-liquid demixing whereas the left-hand figure shows the mechanism for the delayed onset of liquid-liquid demixing*

concept of membrane formation will be described in terms of three components: nonsolvent (1), solvent (2), and polymer (3).

The change in composition may be considered as determined by the diffusion of the solvent (J_2) and of the nonsolvent (J_1) (see Figure 6) in a polymer fixed frame of reference. The fluxes J_1 and J_1 at any point in the cast film can be represented by a phenomenological framework based on thermodynamics of irreversible processes. Here, $\partial\mu_j/\partial x$ - the gradient in the chemical potential - is the driving force for mass transfer of component j at any point in the film and L_{ij} is the permeability coefficient.

Using this approach, one is able to calculate composition paths throughout the thickness of the membranes. It allows to determine whether a polymer solution is immediately unstable and demixes rapidly, or whether diffusional processes proceed for some given time, before delayed demixing occurs. The two different cases are shown schematically in Figure 7.

$$J = -\sum_{j=1}^{2} L_{ij}\left(\phi_i, \phi_j\right)\frac{\partial \mu_j}{\partial x} \qquad (i = 1, 2) \tag{1}$$

It has been an for a long time to be able to predict the development of the morphology as a function of time and even obtain an approximation of the final structure of the membrane. However, such modelling allowed to predict some important trends. When liquid-liquid demixing occurs instantaneously, membranes with a relatively porous top layer are obtained. This demixing

mechanism results in the formation of a porous membrane (microfiltration/ultrafiltration type). However, when liquid-liquid demixing sets in after a finite period of time, membranes with a relatively dense top layer are obtained. This demixing process results in the formation of dense membranes (gas separation/pervaporation). In both cases the thickness of the top layer is dependent on all kind of membrane formation parameters (i.e. polymer concentration, coagulation procedure, additives).

2.4 Integrally skinned membranes

Integrally skinned membranes can be characterised by a defect-free thin toplayer, suitable for gas separation, vapour permeation or pervaporation supported by an open structure. The toplayer has about the same properties as a homogeneous film. The basic requirements of these membranes are similar to composite membranes;

- Toplayer should be thin and absolutely defect-free
- Sublayer should be very open with a negligible resistance

These different structures can be correlated to the two mechanism of membrane formation, viz. a toplayer formed by a delayed onset of demixing, and a sublayer formed by instantaneous demixing. Moreover, a polymer concentration profile should be generated as shown schematically in Figure 3.25, with a high polymer concentration at the top side and a low polymer concentration at the bottom side. Such a profile can be obtained in two ways;

- Introduction of an evaporation step before immersion in a nonsolvent bath (dry-wet phase inversion). As a result of this evaporation step the volatile solvent will evaporate from the surface and a driving force has been generated for diffusion of solvent from the bottom side to the top side. This process may be considered to be convection driven.
- Immersion in a nonsolvent with a low mutual affinity to the solvent (wet phase inversion)

It is possible to achieve a high polymer concentration at the top side by direct immersion in the coagulation bath without any evaporation step. This can be achieved by immersion in a nonsolvent with a low mutual affinity. This results in a high ratio of solvent outflow versus nonsolvent inflow (in fact only the solvent should diffuse out of the polymer film) and a non-linear profile is established as well.

2.5 Simulations of membrane morphology

The above described modelling activities are capable to describe concentrations profiles in the emerging membrane prior to phase separation. However, they can only indicate what type of morphological feature may develope after the phase separation sets in. There are 4 approaches known that aim to address the challenge of predicting morphlogical development during the phase separation process.

1. Cahn-Hillard simulations resolving the emergence of concentration gradient fields in a polymer solution disturbed into a non-equilibrium situation [3]

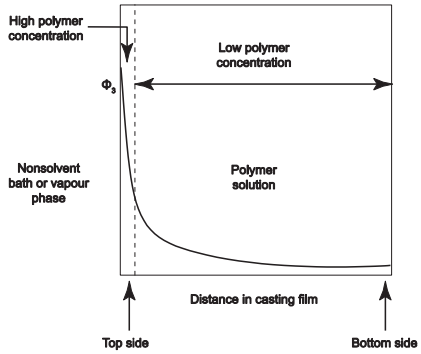


Fig. 8: *Schematic drawing of the volume fraction of polymer (ϕ_3) in the casting solution after a short period of time*

2. Dissipative particle dynamics simulation [4]: the influences of varying the chain length (N) of the polymer composing the membrane, the solvent size in the polymer solution, and the nonsolvent size and amount in the nonsolvent bath on the liquidliquid demixing process and the membrane morphology in detail

3. Stochastic lattice simulations indicating that macrovoids can be avoided by (a) adding non-solvent to the initial polymer solution and (b) high concentrations of a salt that is poorly miscible in the polymer solvent, with no associated decrease in coagulation rate [5].

4. Lattice Bolzmann Simulations which indicate that the influence of polymer solution viscosity and interactions between solvent/nonsolvent/polymer on the simulated structures agree qualitatively with experimental observations [6]

3 Membrane fabrication processes

Porous membranes can be produced on a large scale in two different geometries, (a) as flat sheets up to a width of one meter but infinitely long and (b) as hollow fibers through a continous spinning process. Both processes are highly automated and described in more detail below.

3.1 Flat membranes

The preparation of flat membranes on a semi-technical or technical scale is shown schematically in Figure 9. The polymer is dissolved in a suitable solvent or solvent mixture (which may include additives). The viscosity of the solution depends on the molecular weight of the polymer, its concentration, the kind of solvent (mixture) and the various additives. In Figure 3.45 the polymer solution (often referred to as the casting solution) is cast directly upon a supporting layer, for example a non-woven polyester, by means of a casting knife. The casting thickness can vary roughly from 50 to 500 m. The cast film is then immersed in a nonsolvent bath where exchange occurs between the solvent and nonsolvent and eventually the polymer precipitates. Water is often used (and from an environmental point of view also preferred) as a nonsolvent. But organic solvents (e.g. methanol) can be used as well. Since the solvent/nonsolvent pair is a

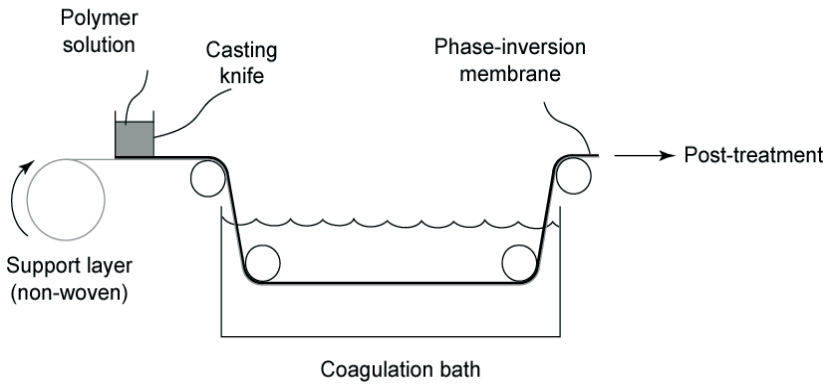


Fig. 9: *Schematic drawing of a dry wet spinning process*

very important parameter in obtaining the desired structure the nonsolvent can not be chosen at will: in fact their compositions are often trade secret of the membrane producers.

3.2 Hollow fibers

Although both flat membranes and hollow fiber membranes can exhibit similar performances, the procedures for their preparation are not the same. Since hollow fibers are self-supporting, the fiber dimensions are very important. Furthermore, demixing takes place from the bore side or lumen and from the shell side or outside, whereas in the preparation of flat membrane demixing occurs from only one side. Spinning parameters are also important with respect to membrane performance during the preparation of hollow fiber. Hollow fibers and capillaries are mostly prepared using a wet spinning or dry-wet spinning spinning process.
A schematic drawing of the dry-wet spinning process is shown in Figure . A viscous polymer solution containing a polymer, solvent and sometimes additives (e.g. a second polymer or a nonsolvent) is pumped through a spinneret, the polymer solution being filtered before it enters the spinneret. The viscosity of the polymer solution must be high (in general more than 100 Poise). The bore injection fluid is pumped through the inner tube of the spinneret. After a short residence time in the air or a controlled atmosphere (the term dry originates from this step) the fiber is immersed in a nonsolvent bath where coagulation occurs. After a washing step, the fiber is then collected upon a godet.
The main spinning parameters are: the extrusion rate of the polymer solution; the bore fluid rate; the 'tearing-rate'; the residence time in the air-gap; and the dimensions of the spinneret. These parameters interact with the membrane-forming parameters such as the composition of the polymer solution, the composition of the coagulation bath, and its temperature.

4 Summary

Polymer membranes often show intricate porous structural features. The use of such features is well established industrially. The control of such porosity features is evidence, however not well understood. Simulations approaches combine thermodynamics of complex mixture with kinetic models to trace diffusion of multiple components and to simulate the emergence of liq-

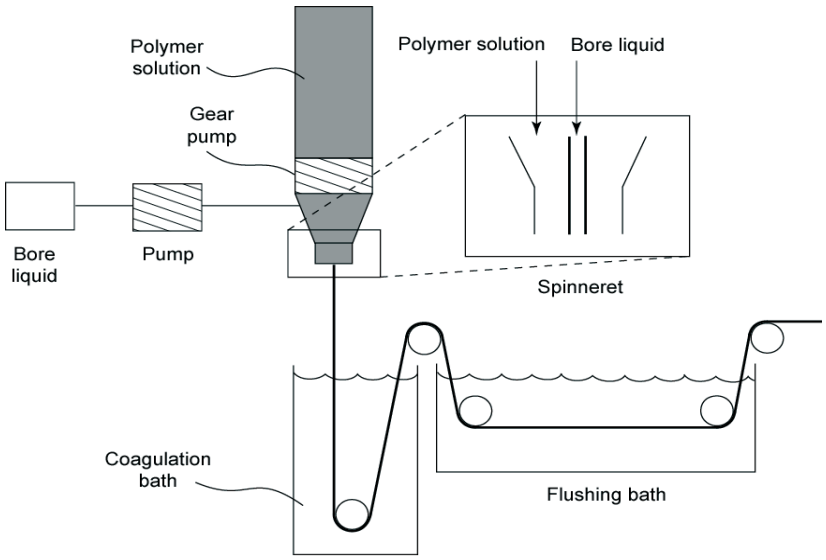


Fig. 10: *Schematic drawing of a dry wet spinning process*

uid and condensed phases. Such simulations are currently at its infant state.

Acknowledgement This chapter is dedicated to belated Prof. M.H.V. Mulder of the University of Twente which I had the pleasure to work with. Following his wish, I have edited his book ”Basic Principle of Membrane Technology”, reworked drawings, rewrote text with coworkers at the University of Twente during the years 2005 and 2006. Unfortunately, the effort never resulted in a new edition of the book as it turned out to be a Herculean task. Parts of this manuscript in fact is edited material stemming from this period.

References

[1] M. Mulder, *Basic principles of membrane technology*.

[2] P. van de Witte, P. J. Dijkstra, J. W. A. van den Berg, and J. Feijen, Journal of Membrane Science **117**(1-2), 1 (1996).

[3] B. F. Barton and A. J. McHugh, Journal of Polymer Science Part B: Polymer Physics **37**(13), 1449 (1999).

[4] X.-L. Wang, H.-J. Qian, L.-J. Chen, Z.-Y. Lu, and Z.-S. Li, Journal of Membrane Science **311**(1-2), 251 (2008).

[5] Y. Termonia, Journal of Membrane Science **104**(1-2), 173 (1995).

[6] A. Akthakul, C. E. Scott, A. M. Mayes, and A. J. Wagner, Journal of Membrane Science **249**(1-2), 213 (2005).

B 6 Polymerdynamics

D. Richter
Institute of Complex Systems, ICS-1
Jülich Centre for Neutron Science, JCNS-1
Forschungszentrum Jülich GmbH

Contents

1 Introduction 2
2 Macroscopic Dynamics 3
3 Neutron Scattering on the dynamics of large scale systems 6
4 The Standard Model of Polymer Dynamics (Rouse) 7
5 Hydrodynamic Interaction in Polymer Solutions 13
5.1 The Zimm Model 13
5.2 Results on Dilute Solutions 15
5.3 Semidilute Solutions 15
6 Long Chains: Entanglements 19
6.1 Mean square displacements 21
6.2 Single chain dynamic structure factor 23
7 Limits of the Reptation Model 25
7.1 Contour Length Fluctuations 26
7.2 Constraint Release 28
8 Dynamics of Ring Polymers 31
9 The Effects of Surfaces and Confinement 36
10 Conclusion and outlook 39
11 References 41

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

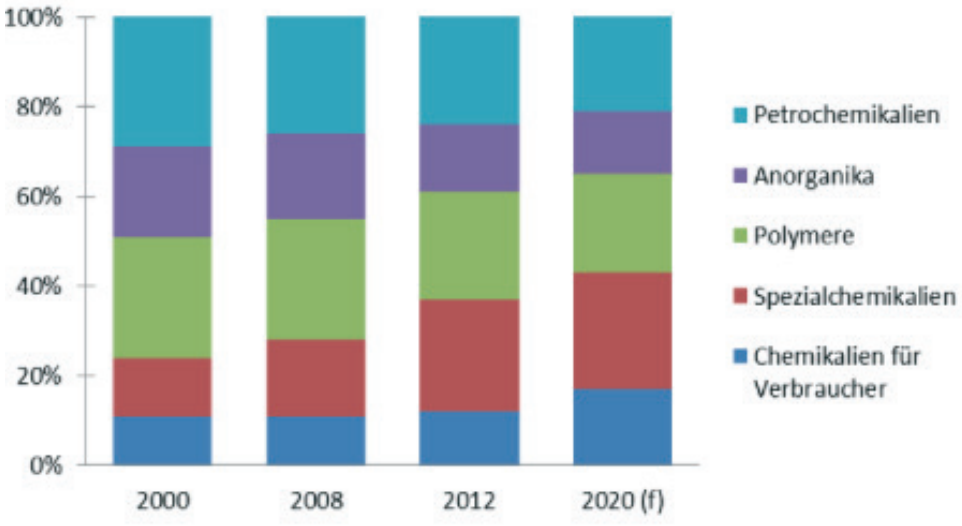
1 Introduction

Look around the house and you will see that you are surrounded by many kinds of polymers: plastic containers, surface coatings in the kitchen, toys and clothes in the living room and bedroom. Modern equipment and components in your car and work place are often made of or coated with polymer composite materials. Today traditional materials, such as metals, ceramics and wood have partly been replaced by synthetic polymers which may be stronger, lighter and cheaper and which through scientific research can be tailored to specific requirements.

It is hard to imagine life without modern synthetic plastics and rubbers. These polymers can be moulded into almost any shape, extruded into thin films and fibres, applied as coatings and given bright colours or made transparent. New polymer composites are continually being developed including reinforced rubbers or construction materials even for aeroplanes.

Polymers are one of the most important products of chemical industry. The development of this industry in Germany started in the second half of the 19th century. BASF was founded in 1866. In 1885 it had already 2335 employees. In 1900 BASF had grown to 6771 employees. Similarly Bayer was founded in 1881. In 1885 Bayer employed 24 chemists and 300 workers. Just 11 years later it had grown to 104 chemists and 2644 workers. The 2012 turnover of chemicals in Europe is given in Table 1. Among these products polymers are on the third rank. Thus, polymers are indeed a very important commodity.

Table 1:
Turnover along products (Europe 2012)



The diffusional motions of long flexible polymers constitute fascinating physics and at the same time represent one of the great challenges of modern material science. The drive towards the molecular understanding of the complex viscoelastic properties of polymer

liquids is the focal point of rheology and connects the classical chemical engineering approach with modern physics [1]. There the tube model invented by Doi and Edwards [2] and de Gennes [3] has shown itself as the most successful molecular model describing the topological confinement imposed by the mutually interpenetrating polymer chains in the melt. In terms of this so called reptation model a theory of viscoelasticity has been developed that describes the main features of polymer melt rheology.

This lecture aims to identify general principles of chain motion on a molecular scale which underpin the macroscopic properties and presents concepts and experimental results on these motional mechanisms in space and time. On the example of linear homopolymers we shall address the basic features of polymer dynamics considering both the dynamics of melts and solutions. Then we discuss two recent topics of polymer dynamics: (i) the unusual dynamics of ring polymers that because of the their unique topology stands out among all other polymer architectures; (ii) the effect of attractive surfaces on the chain dynamics, a topic that is highly important for the understanding of modern composites.

2 Macroscopic Dynamics

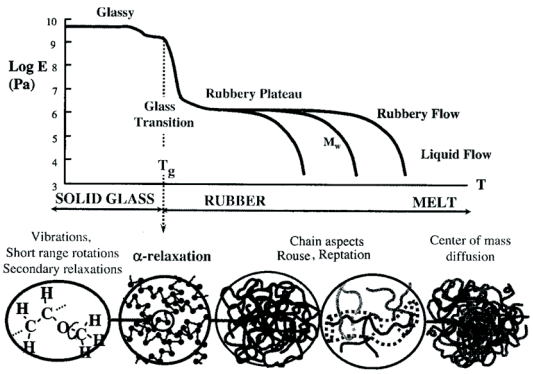


Fig. 1: *Schematic presentation of the modulus of a polymer melt as a function of temperature. The sketches below symbolise the types of motions that give rise to the macroscopic behaviour.*

Dynamic processes in polymers occur over a wide range of length and time scales. Fig. 1 relates the dynamic modulus, as it may be observed on a polymer melt, with the length and time scales of molecular motion underlying the rheological behaviour. Our example deals with an amorphous polymer excluding any crystallization processes. It is clear, that we can distinguish several different regimes. At low temperatures the material is in a glassy state and only small amplitude motions like vibrations, short range rotations or secondary relaxations take place. At the glass transition temperature T_g the primary relaxation (alpha relaxation) becomes active allowing the system to flow. The time range over which this relaxation takes place easily covers more than ten orders of magnitudes in time. The following rubbery plateau in the modulus relates to large scale motions within a polymer chain. Two aspects stand out.

The first is the entropy driven relaxation of out of equilibrium fluctuations, secondly these relaxations are limited by confinement effects caused by the mutually interpenetrating chains. As we shall see later, this confinement is modelled most successfully in terms of the reptation model that was developed by de Gennes. Finally, when the chain has lost the memory of its confined state, liquid flow sets in. That is characterised by the translational centre of mass diffusion of the chain. Depending on the molecular weight, the characteristic length scales from the motion of a single bond to the overall chain diffusion may cover about three orders of magnitude, while the associated time scales may stretch over more than ten orders.

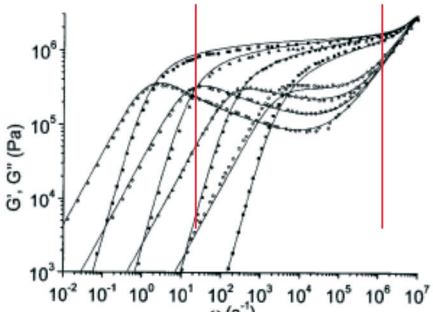


Fig. 2: *Real and imaginary part G' and G'' of the dynamic modulus for polymer melt. The two red lines display the region where the polymer melt responses elastically.*

Fig. 2 quantifies this behaviour on the example of the real and imaginary part of the dynamic modulus which is plotted as a function of frequency covering about ten orders of magnitude. The parameter for the different curves is the molecular weight: the larger the molecular weight, the broader the spectrum of the modulus. Looking on the real part G' we realise a plateau in frequency that enlarges with increasing molecular weight. In this regime the polymer liquid responses elastically like a rubber. Only at low frequencies we see the transition to liquid like flow. Likewise the imaginary part G'', that describes the dissipative behaviour of the melt, exhibits a maximum where the liquid flows sets in. At this point a transition from elastic to liquid like behaviour occurs. The dynamic modulus displays the viscoelastic properties of polymer melts. In a certain frequency range the elastic behaviour prevails while in others we deal with typical liquid like behaviour.

Another characteristic behaviour of polymer melts are universal power laws in the molecular weight dependence of viscosity and diffusion.

Fig. 3 presents the molecular weight or chain length dependence of the melt viscosity for a number of different polymers in a double logarithmic plot. In all cases the viscosity shows two different power law regimes. At low molecular weight the viscosity increases proportional to molecular weight, while above a critical molecular weight M_c, the viscosity increases dramatically with M following a power law with an exponent of about 3.4. Thus, with increasing chain length a polymer melt becomes very tough and viscous.

Likewise, the molecular weight dependence of the translational diffusion coefficient of a polymer in a melt is characterized by two different power law regimes.

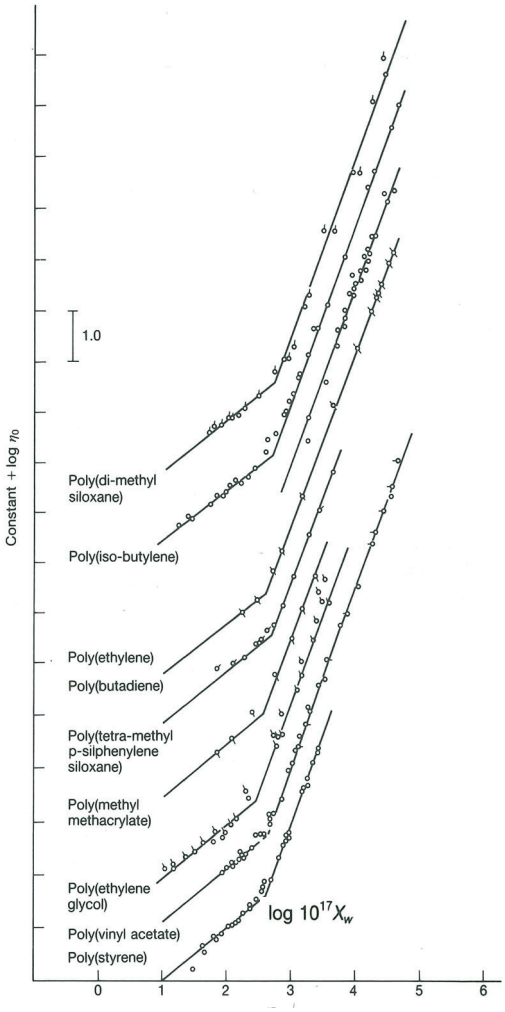


Fig. 3: *Polymer melt viscosity for various polymers as a function of chain length. The data are characterized by cross over between two power laws /4/.*

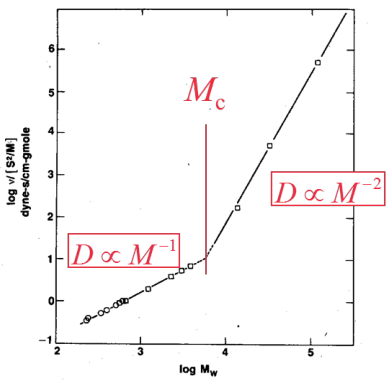


Fig. 4: *Translational diffusion coefficient in a polymer melt measured for polyethylene as a function of molecular weight: Again two different power laws are visible.*

This is shown in Fig. 4, where the inverse diffusion coefficient is displayed as a function of molecular weight in a double logarithmic form. Again, at low molecular weight we realise a linear dependence while at higher molecular weight the diffusion coefficient is inversely proportional to the square of the molecular weight. These earlier results by Persson at al [5]were later modified by more precise experiments reveiling a power law exponent of -2.3 instead of -2 for the molecular weight dependence of the diffusion coefficient [6].

In the following we will now ask for a molecular understanding of this peculiar behaviour of polymer melts. We will go through a hierarchy of models that will let us understand, why long chain molecules exibit the shown dynamical features.

3 Neutron Scattering on the dynamics of large scale systems

Neutron scattering with its space time sensitivity on a molecular and atomic scale unravels the details of the molecular motions in question. Commencing at the scale of the single bond, where movements take place at a pace as in normal liquids, quasielastic neutron scattering (QENS) provides insight into local relaxation processes. At larger length scales first the entropy driven Rouse motion and at even larger distances the effect of entanglement constraints due to the mutual interpenetration of chains comes into the observation range. The most powerful technique suitable for these investigations, the neutron spin echo spectroscopy (NSE) operates in the time domain and uncovers a time range from about *2ps* to *600ns* and accesses momentum transfers between 0.01Å^{-1} and 3Å^{-1}.

Coherent quasi- and inelastic neutron scattering reveals the dynamic structure factor $S(Q,t)$ or its Fourier transformed counterpart $S(Q,\omega)$ [1]

$$S(Q,t) = \frac{1}{N}\sum_{ij}\left\langle \exp\left(-i\underline{Q}\underline{r}_j(t)\right)\exp\left(i\underline{Q}\underline{r}_i(0)\right)\right\rangle \qquad (1)$$

where $\underline{r}_j(t)$ and $\underline{r}_i(0)$ are the position vectors of the scatterers at time t and time $t = 0$ respectively, N is the number of scatterers and $\hbar\underline{Q}/(2\pi)$ is the momentum transfer during scattering. The brackets denote the thermal average. $S(Q,t)$ reflects the pair correlation function and relates to the collective properties of a material. In the neutron cross section it is weighted by the average scattering length $\left|\bar{b}\right|^2$.

Incoherent scattering is related to scattering length disorder which may either origin from spin dependent scattering lengths like in the case of hydrogen or from isotope mixtures with different scattering properties. This disorder prevents constructive interference of partial waves scattered at different atoms and reveals the self correlation function. Eq.[1] provides the self correlation function, if in the double sum only terms with $i = j$ are considered. In the cross section $S_{inc}(Q,t)$ is weighted by the average scattering length fluctuations $\left(\left\langle b^2\right\rangle - \left|\bar{b}^2\right|\right)$.

In Gaussian approximation which is widely used for the calculation of neutron dynamic structure factors for polymer dynamics Eq.[1] is approximated by

$$S_{inc}(Q,t) = \frac{1}{N}\sum \exp\left[-\frac{Q^2}{6}\left\langle \left(r_i(t) - r_i(0)\right)^2 \right\rangle\right] \tag{2}$$

4 The Standard Model of Polymer Dynamics (Rouse)

If we want to describe the motion of a polymer, we could start with the atoms of a chain and solve Newton's equations. This asks us to deal with very many variables - already the simplest polymer chain polyethylene built from CH_2 units, at a reasonable length of about a thousand units features already 3 000 atoms. A melt of such chains gets difficult to treat already for advanced molecular dynamics simulations. We may make a step further and coarse grain in a way, that we describe the atoms along one bond, in this case the CH_2 unit by one entity leading to the unified atom model. In this case, we still have thousand atoms in one chain. Again we need severe MD simulation in order to solve the problem and we still don't have a model. In order to go further, we have to coarse grain significantly more and still keep the essentials of the problem (Fig. 5).

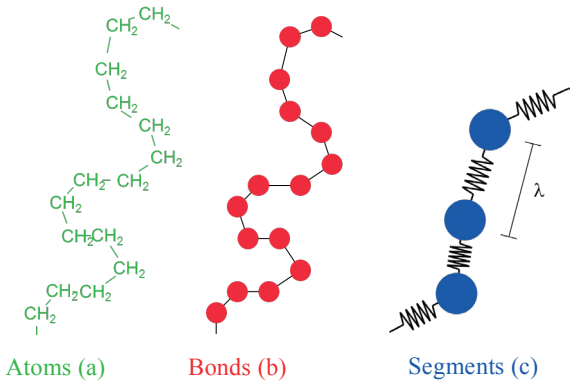


Fig. 5: *Schematic representation of the coarse graining process that leads to the Rouse model.*

a) Polyethylene chain
b) Unified atom chain
c) Chain represented by beads and springs

This is achieved with the Rouse model [7]. Here the polymer chain is described by a sequence of beads and springs where the beads undergo friction with a heat bath. The springs originate from the chain entropy that prefers a Gaussian chain conformation. Any deviations from such conformations undergo a restoring force of harmonic character.

With that we can write down a Langevin equation for the chain segmental motion.

$$\zeta_0 \frac{dr_n}{dt} = \frac{k_B T}{l^2}(r_{n+1} - 2r_n + r_{n-1}) + f_n(t) \tag{3a}$$

$$\zeta_0 \frac{dr_n}{dt} = k\frac{\partial^2 r_n}{\partial n^2} + f_n(t) \tag{3b}$$

Here ζ_0 is the friction of the beat with the heat bath $k_BT/l^2 \equiv k$ is the entropic spring constant, where l is the segment length and $f_n(t)$ describes the thermal random force acting on bead "n". Assuming white noise this equation can be solved exactly (see e.g./3/). Equ.[3b] is the

continuous version of Equ.[3a], where the difference term is replaced by the second derivative with respect to the now continuous monomer index "n". For the viscosity of the melt we get

$$\eta = \frac{\zeta_0 l^2 N \rho}{36} \tag{4}$$

where ρ is the density and N the chain length. Note, that this viscosity is proportional to the number of chain segments N.

Similarly for the translational centre of mass diffusion coefficient Eq. [3] leads to.

$$D = \frac{k_B T}{\zeta_o N} \propto N^{-1} \tag{5}$$

The diffusion coefficient is inversely proportional to the number of friction exerting beads. For short chains both results agree with macroscopic experiments, but we may ask is this model also correct microscopically?
For this purpose we have to look on the chain motion on the scale of the chain. There a prominent quantity is the mean square segment displacement. For long times we know that the motion has to be diffusive and therefore $< \Delta r_n^2(t) >= 6Dt$.

But what happens, when the chain segments have moved distances smaller than the chain dimensions? Let's start with a borderline case, the case where the covered distance is just the chain size. This obviously is the longest time, where internal correlations within the chain could play a role. This longest relaxation time is also called the Rouse time τ_R. From the simple consideration we would get $6D\tau = R_E^2 = Nl^2$, where R_E is the chain end to end distance.

Using Eq [5] and equating for τ_R we get

$$\tau_R \simeq \frac{N^2 l^2 \zeta_0}{6 k_B T} \tag{6}$$

differing only slightly from the correct value. A solution of Eq.[3] yields
$\tau_R = \frac{N^2 l^2 \zeta_0}{3\pi^2 k_B T} = \tau_0 N^2; \tau_0 = \frac{l^2 \zeta_0}{3\pi^2 k_B T} = \frac{1}{W\pi^2}$. Now, any subsection of the chain with $\frac{N}{p}$ segments relaxes as the whole chain. Therefore, we have

$$\tau_p \propto \tau_0 \left(\frac{N}{p}\right)^2 \tag{7}$$

at the time τ_p the chain section with N/p monomers moves over its own distance. Thus,

$$< (r_n(\tau_p) - r_n(0))^2 >= l^2 \frac{N}{p} = l^2 \left(\frac{\tau_p}{\tau_0} \right)^{\frac{1}{2}} \tag{8}$$

the most right part of Eq. [8] follows from Eq. [7].

In a more formal way this behaviour may derived in terms of eigenmodes of the chain that exhibit a wavelength $\lambda = l * N / p$ along the chain. These eigenmodes are obtained by a Fourier transformation of the Langevin Equ. [3] with the proper boundary conditions of force free ends (see ref./4/). They turn out as

$$\phi_p(n) = \frac{1}{N} \cos\left(\frac{p\pi}{N} n \right) \tag{9}$$

These modes relax with the characteristic times $\tau_p = \tau_R / p^2$.

Since this Equ.[8] holds for all p, the chain segments at times shorter than τ_R move in a subdiffusive way. The mean square displacement only increases with the square root of time. This is a basic prediction of the Rouse model. Performing the full calculation starting from Eq. [3], the final result for the time dependent mean square displacement is:

$$r_n^2(t) >= \sqrt{\frac{12 l^2 k_B T}{\pi \zeta_0}} t \tag{10}$$

The segment self correlation function that is measured with quasielastic incoherent neutron scattering directly accesses this quantity. In Gaussian approximation we have

$$S_{self}(Q,t) = \exp\left[-\frac{Q^2}{6} < r^2(t) > \right] = \exp[-D_R Q^2 t] * \exp\left[-\frac{2}{\sqrt{\pi}} (\Omega_R(Q) t)^{\frac{1}{2}} \right]; \Omega_R(Q) = \frac{k_B T l^2}{12 \zeta_0} Q^4 \tag{11}$$

The second part of this equation is obtained by inserting Eq. [10]. $\Omega_R(Q)$ is the characteristic relaxation rate, that increases with the momentum transfer Q^4.

Even though a clear cut prediction, experimentally the observation of the self correlation function of a Rouse chain is an important challenge. The necessary resolution at the low momentum transfers requires, neutron spin echo spectroscopy [8]. Here, incoherent experiments are difficult, since incoherent scattering depolarises the beam to a large extend (2/3 spin flip scattering). Therefore, using a trick the first successful experiments were carried out. The chemists produced deuterated PDMS where randomly short protonated sections were copolymerised. These protonated sections in a generally deuterated environment gave rise to coherent scattering, however, since the scattering from different labels was uncorrelated the self correlation function was measured.

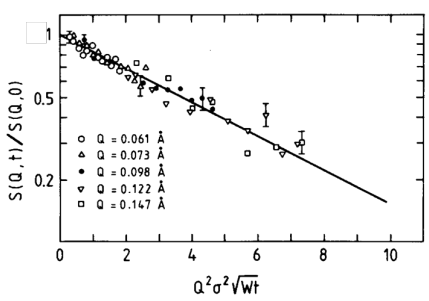


Fig. 6: *Dynamic structure factor for the segmental self motion in a PDMS-melt. The data are scaled with the Rouse variable. Solid line predicted* $\sqrt{t}$ *relaxation by the Rouse model* $\sigma^2 \equiv l^2$.

Fig. 6 displays the obtained self correlation function for PDMS [9] in a presentation where the logarithm of the scattering function is plotted versus $(\Omega_R(Q)*t)^{1/2}$ the scaling variable of Eq. [11]. In this way all the data collapse on one single master curve that according to Eq. [11] should be a straight line. The experimental results beautifully verify the major prediction of the Rouse model and show that the simple approximation of the bead - spring model properly accounts for the segmental dynamics of the PDMS chain on the space time frame investigated.

For the single chain dynamic structure factor, where we look on a labelled e.g. protonated chain in a deuterated environment, we have to deal with the interference of scattered waves originating from the different atoms or monomers of the chain. The detailed calculations are found in reference /3/. The result may be expressed in terms of the Rouse modes (Equ.[9]) and the corresponding relaxation times $\tau_p = \tau_R / p^2$.

$$S_{chain}(Q,t) = \frac{1}{N}\exp\left[-Q^2 D_R t\right]\sum_{i,j}\exp\left\{-\frac{1}{6}|i-j|Q^2 l^2\right\}* \\ \exp\left\{-\frac{2}{3}\frac{R_E^2 Q^2}{\pi^2}\sum_p \frac{1}{p^2}\cos\left(\frac{p\pi i}{N}\right)\cos\left(\frac{p\pi j}{N}\right)\left(1-\exp\left(-\frac{tp^2}{\tau_R}\right)\right)\right\} \tag{12}$$

for small Q ($QR_E < 1$) the second and third terms are negligible and $S_{chain}(Q,t)$ describes the centre of mass diffusion of the chain.

$$S_{self}(Q,t) = \frac{1}{N} S_{chain}(Q,t) = \exp\left(-D_R Q^2 t\right) \tag{13}$$

For $QR_E > 1$ and $t < \tau_R$ the internal relaxations dominate. For $t = 0$ we have $S_{chain}(Q,t) = S_{chain}(Q)$; i.e. the structure factor corresponds to a snapshot of the chain structure.

$$S_{chain}(Q) = \frac{1}{N}\sum_{n,m}\exp\left(-\frac{1}{6}Q^2|n-m|^2\right) \tag{14}$$

Replacing the summations by integrals and observing the relation $R_g^2 = \frac{1}{6} N l^2$ for the radius of gyration Eq.[14] immediately leads to the well known Debye function

$$S_{chain}(Q) = N f_{Debye}\left(Q^2 R_g^2\right)$$
$$f_{Debye}(x) = \frac{2}{x^2}\left(e^{-x} - 1 + x\right) \qquad (15)$$

Similar to the self correlation function chain still follows the universal decay if plotted versus the Rouse variable $(\Omega_R * t)^{1/2}$.

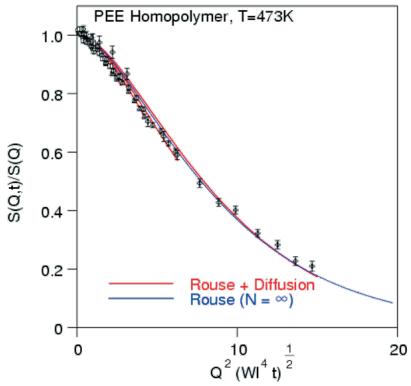


Fig. 7: *Single chain dynamic structure factor measured for polyethylethylene scaled with the Rouse variable. All data collapse to a single master curve (see text).*

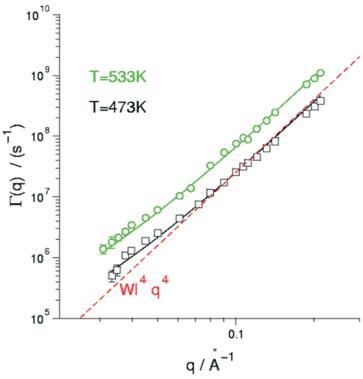


Fig. 8: *Characteristic relaxation rate for polyethylethylene at two different temperatures as a function of momentum transfer Q. The data display the predicted Q^4 scaling. For the deviations at low Q see text.*

This is displayed for polyethylethylene (PEE) in Fig. 7 [10] again the data follow closely a single master curve, however, with a shape that differs from the simple form of Fig. 6. Furthermore, again the characteristic rate Ω_R is predicted to follow the Q^4 power law.
In Fig. 8 this is beautifully seen, where the characteristic width $\Gamma(Q) \equiv \Omega$ is displayed as a function of Q in a double logarithmic way. For more than three orders of magnitude in rate, the data display the Q^4 law. The small deviations at low Q result from the translational diffusion coefficient, that additionally contributes to the overall relaxation. From Eq. [11] follows

$$\Gamma(Q) = D_R Q^2 + \Omega_R(Q) = Q^2 (D_R + \frac{k_B T l^2 Q^2}{12\xi_0}) \qquad (16)$$

The solid lines in Fig. 8 reflect the combined diffusive and internal relaxation behaviour.

The above expressions provide an universal description of the dynamics of a Gaussian chain and are valid for real linear polymer chains on intermediate length scale. The specific (chemical) properties of a polymer enter only in terms of two parameters $N\mathrm{l}^2 = R_E^2$ and ℓ^2/ζ_0. The friction parameter is governing the Rouse variable in terms of $W\mathrm{l}^4 = \frac{3k_BT\mathrm{l}^2}{\zeta_0}$. As eluded to in Eq.[16] also the center of mass diffusion coefficient may be expressed in these terms $D = k_BT\mathrm{l}^2 \Big/ \left(\zeta_0 R_E^2\right)$. Since the Rouse model does not contain an inherent length scale, the parameter N (chain length) and ℓ^2 (segment length squared) are somewhat arbitrary as long as the physical values of ℓ^2/ζ_0 and R_E^2 are kept constant. The NSE experiments measure directly the friction coefficient/length squared.

Setting the time scale the monomeric friction coefficient is a basic quantity in all rheological measurements. There this quantity is inferred indirectly either from measurements of the dynamic modulus $G(\omega)$ in the first relaxation regime. Here in terms of the Rouse model the relaxation spectrum H is given by [11]

$$H(\omega) = \frac{\sqrt{\langle l^2 \rangle}\rho N_A}{2\pi M_0}\left(\frac{\zeta_0 k_B T}{6}\right)\omega^{1/2} \tag{17}$$

M_0 is the monomer mass, N_A the Avogadro number, ζ_0 the friction coefficient related to the monomer size and ω the relaxation frequency; or from viscosity measurements (Equ.[4]). Very often also viscosity studies at high molecular weight are used. Then the effect of entanglements has to be considered in an empirical fashion. The NSE experiments in the Rouse regime provide direct microscopic access to ζ_0. From $W\mathrm{l}^4 = \frac{3k_BT\mathrm{l}^2}{\zeta_0}$ we obtain

$$\zeta_0 = \left(\frac{3k_BT}{Wl^4}\right)l^2 \tag{18}$$

A detailed tabulation is e.g. given in ref. [12]. NSE experiments on polymer melts in the Rouse regime have been performed by now on seven different polymers including two studies of a systematic temperature dependence [1] Table 1 presents the monomeric friction coefficient originating from these measurements and compares them with rheological data taken mostly from Ferry's book. While in some instances very good agreement is evident (PDMS, PVE) in other cases substantial deviations are visible. Also the method, how the rheological results were obtained, seems to determine significantly the rheological friction coefficient e.g. in the case PEP a rescaling of the high molecular weight viscosity at 373K gives $\zeta^{VIS} = 6.3\cdot10^{-11}\,{}^{Ns}/_{m}$ while an evaluation from the dynamic modulus yields $\zeta^{Mod} = 0.98\cdot10^{-11}\,{}^{Ns}/_{m}$. The microscopic value of $3.8\cdot10^{-9}\,{}^{Ns}/_{m}$ is just in between.

In summary the chain dynamics at intermediate scales is excellently described by the bead-spring Rouse model:

- it predicts correctly, the Q-dependence of the relaxation rate (Rouse scaling variable)

- it provides quantitative expressions for the spectral shape of self and pair correlation functions
- it furthermore establishes the correct reference to macroscopic quantities as viscosity ζ and translational diffusion coefficients D_R

Table 1:
Monomeric friction coefficients from NSE experiments compared to rheological data from the compilation of Ferry [11].

Polymer	Temperature	$\zeta^{NSE}\left[10^{11}\frac{Ns}{m}\right]$	$\zeta^{rheo}\left[10^{11}\frac{Ns}{m}\right]$
PVE [10]	408	1.59	1.26
PI [10]	408	0.569	0.19
1,4 PB [11]	353	0.60	3.1
PEP [12,13]	373	3.82	0.98 (6.3*)
PE [12,13]	415	0.18	0.255*
PDMS [14]	373	0.298	0.282
PEE [7]	473	0.695	

*)J. Luettmer-Strathmann, Int. J. Thermophys. **22**, 1507 (2001)

5 Hydrodynamic Interaction in Polymer Solutions

In this chapter we will discuss a phenomenon occurring in solutions only. As discussed broadly in the lecture on "Hydrodynamics", in solutions, hydrodynamic interactions transmitted by the solvent are an important ingredient for the understanding of the dynamics of such systems. If a monomer moves, via friction it exerts a force on the solvent creating a flow field. This flow field then interacts again via friction with the other polymer segments and couples the motions of all monomers with each other. In this paragraph we will first discuss the conceptual basis in terms of the Oseen tensor and the Zimm model, then we will present experimental results on dilute solutions, discuss the effect of concentration in regarding semidilute solutions, where the screening of the hydrodynamic interactions and collective diffusion are important.

5.1 The Zimm Model

The flow field $\underline{v}$ exerted by a force $\underline{F}$ in the solvent is described by the Oseen tensor $\underline{\underline{T}}$. $\underline{\underline{T}}$ is derived from the imcompressible Navier-Stokes equation.

$$\underline{v}(\underline{r}) = \underline{\underline{T}} \cdot \underline{F} \tag{19}$$

$$\underline{\underline{T}}(r) = \frac{1}{8\pi\eta_s|r|}\left\{\underline{\underline{E}} - \frac{\underline{r}\otimes\underline{r}}{r^2}\right\} \tag{20}$$

η_s is the solvent viscosity, and r the distance between force and flow field. $\underline{\underline{E}}$ is the unity tensor and the circle with the cross designates the outer product.

For a polymer in solution the velocity of the solvent $\underline{v}(\underline{r}_n)$ at the monomer position $\underline{r}_n$ is the consequence of the forces which are exerted by all other moving monomers on the solvent.

$$\underline{v}\ \left(\underline{r}_n\right) = \sum_{m \neq n}\ T_{nm}\quad \underline{F}_m^{chain} \tag{21}$$

$\underline{F}_m^{chain}$ is the force which is exerted by segment m on the solvent. This force is given by the Rouse equation (Eq. [3b]). Inserting Eq.[3b] into Eq.[21], we obtain

$$\underline{v}(\underline{r}_n) = \sum_{m \neq n}\ T_{nm}\ \left(k\frac{\partial^2 \underline{r}_m}{\partial m^2} + \underline{f}_m \right) \tag{22}$$

This velocity field exerts a force on segment n. Including the hydrodynamic force term in the Rouse Eq. [3b], we obtain the Zimm equation [23].

$$\zeta \frac{d\,\underline{r}_n}{dt} = \left(k\ \frac{\partial^2 \underline{r}_n}{\partial n^2} + \underline{f}_n \right) + \zeta \sum_{m \neq n} \underline{\underline{T}}_{nm}\ \left(k\frac{\partial^2 \underline{r}_m}{\partial m^2} + \underline{f}_m \right) \tag{23}$$

This is a coupled non-linear differential equation which cannot be solved without approximations.

The standard solution is obtained after an orientational averaging of the Oseen tensor. Furthermore, all intrachain distances are averaged with their equilibrium distribution. For a theta solvent this equilibrium distribution is Gaussian. With that, the Oseen tensor becomes

$$\left\langle T_{nm} \right\rangle = \frac{1}{\left(6\pi^2 \left| n-m \right|\right)^{1/2} l\eta_s} \tag{24}$$

With these approximations Eq. [23] may be linearized by Fourier transformation using the Rouse normal modes (Eq.[9]) that approximately diagonalize the Zimm equation. Neglecting the very small off-diagonal elements the Zimm equation may be easily solved. For the relaxation time spectrum, we obtain

$$\frac{1}{\tau_p} = \frac{1}{\tau_Z}\, p^{3/2} \text{ with } \quad \tau = \frac{\eta_s N^{3/2} l^3}{\sqrt{3}\pi k_B T} = 0.325\,\frac{\eta_s R_E^3}{k_B T} \tag{25}$$

This different power law in p leads to a Q-scaling of the characteristic relaxation rate $\Omega_Z \sim Q^3$ rather than $\Omega_R \sim Q^4$. Compared to the Rouse model, the relaxation time spectrum is changed qualitatively ($1/\tau_p \approx p^{3/2}$ instead of $1/\tau_p \approx p^2$ in the Rouse case). This qualitative change is the consequence of the long range hydrodynamic interaction. For the mean square displacement of a single monomer again a qualitatively different behaviour compared to the Rouse case is found.

$$\left\langle \Delta r_n^2(t) \right\rangle = \Gamma\frac{1}{3}\frac{2}{\pi^2}\left(\frac{\sqrt{3}\pi k_B T}{\eta_s}t\right)^{2/3} \tag{26}$$

For the diffusion coefficient of the whole chain

$$D_z = \frac{8k_B T}{3\left(6\pi^3\right)^{1/2}\eta_s \mathrm{l}\sqrt{N}} = 0.196\,(0.203)\,\frac{k_B T}{\eta_s \mathrm{R_e}} \tag{27}$$

is obtained, where the first prefactor stands for θ- and the second for good solvents. Polymer coils exhibit a diffusion coefficient following Einstein's law for hard spheres! Table 2 compares the predictions of the Rouse and the Zimm models.

Table 2: *Comparison of predictions of the Rouse and the Zimm models*

Properties	**Zimm**	**Rouse**
Diffusion coefficient	$0.196\frac{k_B T}{\eta_s R_E}$	$\frac{k_B T}{N\zeta_0}$
Longest relaxation time	$\tau_Z = 0.325\frac{\eta_s R_E^3}{k_B T}$	$\tau_R = \frac{\zeta_0 R_E^4}{3\pi^2 k_B T \mathrm{l}^2}$
Dispersion relation	$\frac{1}{\tau_p} \approx \frac{1}{p^{3/2}}$	$\frac{1}{\tau_p} \approx \frac{1}{p^2}$
Mean-square segment displacement	$\Gamma\left(\frac{1}{3}\right)\frac{2}{\pi^2}\left(\frac{\sqrt{3}\pi k_B T}{\eta_s}t\right)^{2/3}$	$2\ell^2\left(\frac{3k_B T}{\pi\zeta\ell^2}t\right)^{1/2}$

5.2 Results on Dilute Solutions

For polydimethylsiloxane (PDMS) bromobenzene is a theta solvent at $T = 84°C$. In such a solvent the polymer-polymer and the polymer solvent interactions compensate such that like in the melt the chain conformations follow an ideal Gaussian random walk. Fig. 9 presents NSE results under such conditions. The solid lines display the prediction of the Zimm model. For the data description the only parameter of the model, the solvent viscosity, was fitted. The resulting value agrees within 10% with that for the solvent viscosity [13].

5.3 Semidilute Solutions

In semidilute solutions the polymer chains start to interpenetrate each other and to overlap. On short scales corresponding to distances smaller than the average distance ξ_c between points of overlap - we expect that the polymer will behave as in dilute solution. For larger distances $r > \xi_c$, we need to consider two different effects.

1. The transient network formed by the overlapping polymers will relax with a collective diffusion coefficient D_{coop}.

2. Since the flow field scatters at the mesh of chains, at such distances the hydrodynamic interactions will be screened.

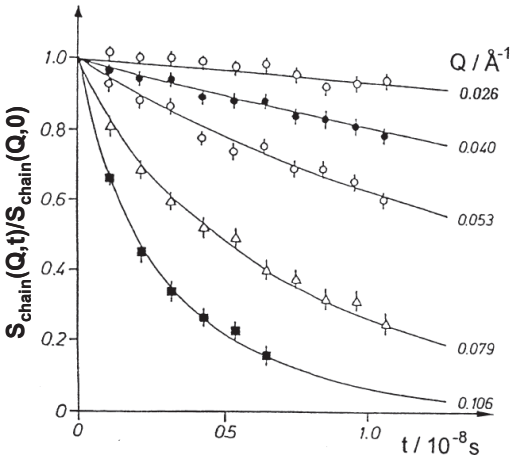


Fig. 9: *Dynamic structure factor obtained from a dilute solution of PDMS in bronobenzene at T = 84°C under θ conditions.*

We will discuss first the collective diffusion. The transition from the spatial heterogeneity at small distances towards homogenous behaviour at $r = \xi_c$ changes the dynamic properties.
At short distances we will observe single chain dynamics with the characteristic relaxation rate, $1/\tau(Q) \approx T/\eta_s \cdot Q^3$ while at larger distances collective relaxation with $1/\tau(Q) \approx D_{\text{coop}} \cdot Q^2$ should prevail. The scaling theory proposed by de Gennes [15] predicts that the collective diffusion in good solvents should be proportional to the polymer volume fraction $\phi^{3/4}$. Analogously, the correlation length should scale with $\xi_c \approx \ell \phi^{-3/4}$.

Fig. 10 displays the reduced relaxation rates $\Omega_Z(Q)/Q^2$ for varying volume fractions of PDMS in benzene as a function of Q. In this presentation the Zimm characteristic rate $\Omega_Z(Q) \sim Q^3$ becomes a straight line, while the collective diffusion changes into Q independent plateaus. We observe, that for the two lowest polymer volume fractions 2% and 5% for the whole momentum transfer regime, single chain behaviour is observed. With increasing polymer concentration, at increasing Q we observe a transition to a plateau characterized by the collective diffusion coefficient. Let us consider the points at $\phi = 0.18$. Here the first four data points ($Q < 0.05$Å^{-1}) define the plateau of the collective diffusion coefficient, while the data points at higher momentum transfer agree with those of the dilute polymer solution. In this Q regime single chain dynamics is observed. The such determined collective diffusion coefficient exhibits a power law with an exponent 0.67 ± 0.05 slightly smaller than the value of 0.75 predicted by the scaling theory [13].

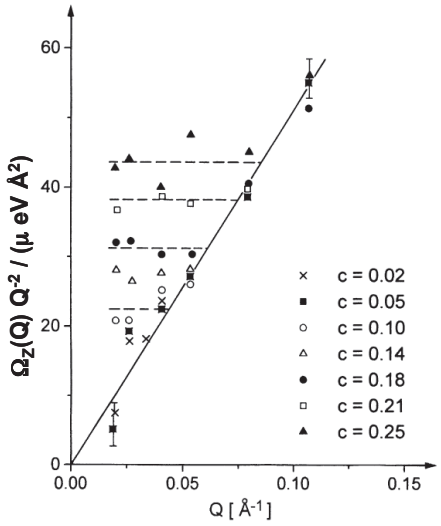


Fig. 10: *Collective reduced relaxation rate $\Omega_Z(Q) \equiv 1/\tau(Q)$ for a PDMS ensemble dissolved in benzene at various volume fractions $\phi \equiv c$. The solid line represents the scaling $\Omega_Z(Q) \sim Q^3$, while the dashed lines denote the Q-dispersion of the collective diffusion dilute solution.*

While the collective behaviour of the polymer ensemble is characterized by the collective diffusion, the relaxation of a single chain is dominated by the screening of the hydrodynamic interaction. While at short distances $r < \xi_c$ the unperturbed hydrodynamic flow field leads to Zimm relaxation, for distances $r > \xi_c$ the hydrodynamic interaction is screened and Rouse relaxation is expected.

An experimental verification of this prediction requires the observation of the dynamics of a single polymer chain within the ensemble. This may be achieved either by labelling a few chains within an ensemble which is matched by the solvent or by experiments at zero average contrast [16]. In this case a mixture of the same amount of identical protonated and deuterated polymers in a solvent with the exact average contrast is investigated. Then, all interference terms between different chains in the structure factor compensate and only the single chain dynamic structure factor is visible (the interference terms between congeneric polymers appear with positive, those between different polymers with negative sign. Since the polymers are identical and the solvent scattering power is exactly in between that of the polymers all interference terms compensate.

Fig. 11 presents the static structure factor of a semidilute solution (i) under average contrast zero (Iz: 50% h/50% d-polymer) and (ii) of 100% protonated polymers in a deuterated solvent. While in the case (i) the single chain structure factor is observed in the case (ii) that of the transient polymer network is seen. There only the polymer parts in between overlap points, the so called blobs, scatter leading to a very much reduced intensity, at low Q.

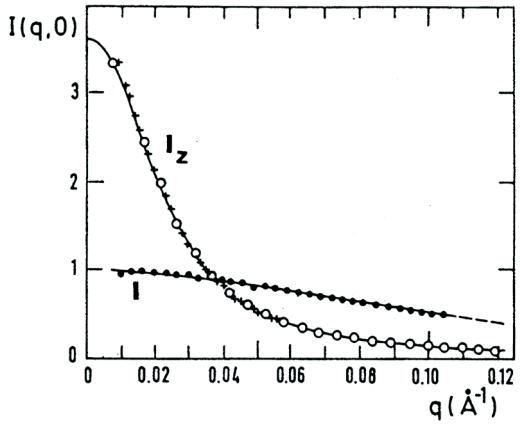


Fig. 11: *Static structure factor of PDMS solutions under full (I) and zero average contrast (I_z) (PDMS concentration 0.2g/cm^3).*

Fig. 12 compares experiments under full contrast with others at zero average contrast above and below the crossover condition $Q\xi_c = 1$. For $Q\xi_c < 1$ in the region of the screened hydrodynamic interaction the experiment under full contrast (100% hPDMS) exhibits a single exponential decay (solid line). In this regime, collective diffusion is observed. The single chain structure factor (50% hPDMS, 50% dPDMS) under zero average contrast displays the square root behaviour of the Rouse dynamics. In this regime, the hydrodynamic interaction is screened and only the Rouse dynamic remains. Going to Q values above the cross over ($Q\xi_c > 1$) under both contrast the $t^{2/3}$ time dependence of the Zimm model is approximately observed. In both cases, the single chain dynamic dominates.

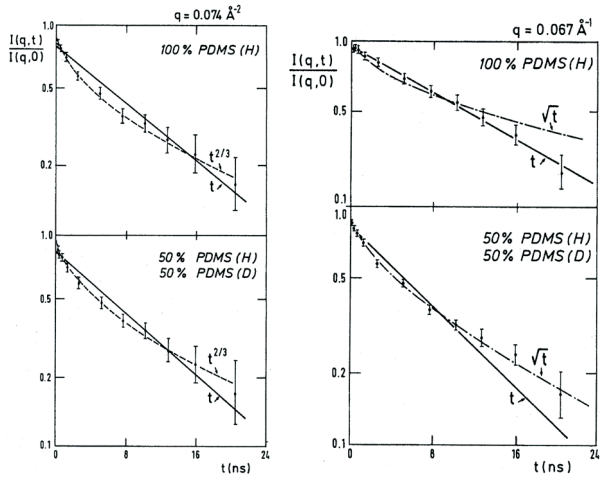


Fig. 12: *Comparison of the dynamic structure factors of semidilute PDMS-solutions under full (100% PDMS) or zero average contrast (50% PDMS (H), 50% PDMS (D)) above and below the cross over $Q\xi_c = 1$.*

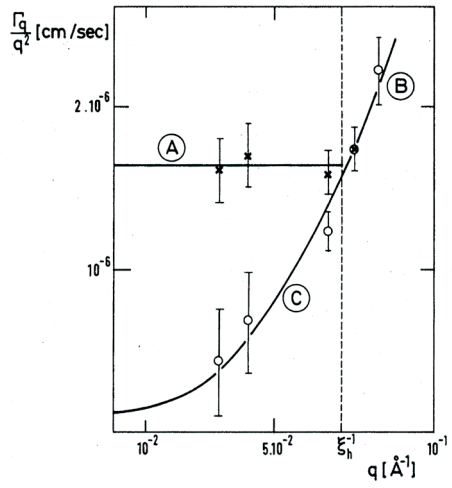


Fig. 13: *Dispersion relation of the characteristic relaxation processes (A) collective diffusion (B) single chain Zimm dynamics (C) Rouse relaxation with screened hydrodynamic interaction.*

Finally, Fig. 13 presents the reduced relaxation rates $\Gamma(Q)/Q^2$. Within regime *B* Zimm dynamic prevails. This is valid for both labelings. On the other hand in the regime of small momentum transfers dependent on the contrast, either the plateau of the collective diffusion (A) or the strong reduction of the relaxation rate as a consequence of the Rouse dynamics is seen (C). These experiments are a particularly beautiful example for the opportunities offered by the h-d contrast variation in neutron scattering.

6 Long Chains: Entanglements

Macroscopically the dynamics of long chain polymer melts is characterised by a plateau regime in the dynamic modulus. Thus, there is a frequency or time regime where a polymer melt responses elastically like a rubber. There, the elastic properties are derived from the entropy elasticity of the chains between permanent cross links. The modulus of a rubber is inversely proportional to the mesh size and proportional to the temperature. In analogy it is suggestive to assume that in a polymer melt entanglements or topological interactions between chains take the role of the rubber cross links. They are supposed to form a temporary network, which displays the rubber elastic properties. However, other than in a rubber, for long times the chains may disentangle and the melt flows. Therefore, the dynamic modulus decays for long times or low frequencies. Using the analogy to the modulus of a rubber, we may estimate the distance between entanglement points from the value of the modulus associated to the plateau, the plateau modulus G_N^0. For different polymers these distances come out to be between about 30 and 100 Å. On that basis a number of theories for viscoelasticity have been developed. The most famous of them is the reptation model by de Gennes [3] and Doi and Edwards [2]. In this model the dominating chain motion is the reptile like creep along the chain profile. The lateral restrictions by the interpenetrating chains are modelled by a tube of size "d" parallel to the coarse grained chain profile. According to theory d relates to the plateau modulus of the melt.

$$d^2 = \frac{4}{5}\frac{R_E^2}{M}\frac{k_B T}{G_N^O} \tag{28}$$

Where R_E is the chain end to end distance and M its mass. The width of the tube or the distance in between entanglements is a mesoscopic quantity significantly larger than the distance between the chain back bones. Thus, there is lateral freedom on intermediate scales and only large scale motion is affected by the tube constraints.

These lateral constraints were first visualized in a famous computer simulation approach by Kremer and Grest [11] in 1990.

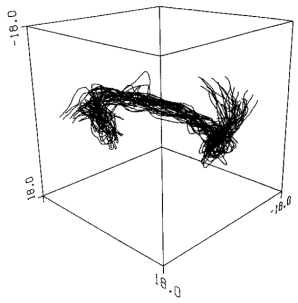


Fig. 14: *Results for the chain contour from a computer simulation by Kremer and Grest [11]. The different lines represent the chain contours at different times.*

Fig. 14 displays their results. What is seen are contour plots of a polymer chain in a melt at different times. We clearly see, that in the middle part of the polymer the chain trajectory stays constrained in a tube like fashion. On both ends however, the chains are able to leave the constraints and explore new areas in space. This is the essence of the tube model.

The tube model also makes the main macroscopic finding for the molecular weight dependence of viscosity and translational diffusion comprehensible. The viscosity relates to the longest relaxation time in a system. If we consider Rouse diffusion along the tube with a Rouse diffusion coefficient $D_R \approx 1/(N\zeta_0)$ then an initial tube configuration is completely forgotten when the mean square displacement along the tube $\langle r^2(t)\rangle_{\text{tube}} = (\text{contour length } L)^2$. Thus, for the longest relaxation time we obtain

$$\tau_\eta \approx \frac{L^2}{D_R} \approx N^3 \zeta_0 \tag{29}$$

The diffusion coefficient is found by considering that during this time in real space the MSD just amounts to the end to end distance of the chain squared. Thus, we obtain

$$D_{rep} = \frac{R_E^2}{\tau_\eta} \approx \frac{N}{N^3 \zeta_0} \approx \frac{1}{N^2 \zeta_0} \tag{30}$$

6.1 Mean square displacements

We now consider the predictions of the reptation model for the mean square displacements of the chain segments [4,15]. For short times, when the chain segments have not yet realized the topological constraints ($r^2 < d^2$) we expect unrestricted Rouse motion $\langle r^2(t)\rangle \approx t^{1/2}$ (Eq.[7]). Experimentally this was the case for PDMS (Fig. 6). At a time $\tau_e = \frac{\pi}{4}\tau_o N_e^{\,2}$, where N_e is the length of an entanglement strand, the mean square displacement reaches the order of the tube diameter. τ_e is derived as the Rouse time for a polymer strand, spanning the tube. Thereafter motional restrictions are expected.

For times $t > \tau_e$ one dimensional curve linear Rouse motion along the tube needs to be considered. Displacements along the tube are described by Eq.[10] where we have to change real space coordinates to coordinates $s(t)$ along the tube. If a segment is displaced along the tube by $\langle (s_n(t) - s_n(0))^2\rangle$ then the mean square displacement in 3-d real space is $d\left\langle \left(s_n(t) - s_n(0)\right)^2 \right\rangle^{1/2}$. With that we obtain

$$\left\langle r^2(t)\right\rangle = \begin{cases} 2d\left(\dfrac{k_B T l^2 t}{\zeta_0 \pi}\right)^{1/4} & \tau_e < t < \tau_R \\ \\ 2d\left(\dfrac{k_B T t}{N\zeta_0}\right)^{1/2} & \tau_R < t < \tau_d \end{cases} \tag{31}$$

In Fig. 15 the two situations correspond to the second and the third process.

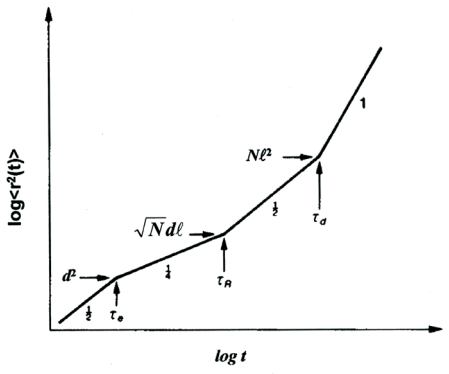


Fig. 15: *Predicted time dependence of the segmental mean square displacements. Several power laws are visible. At short times we observe the Rouse regime (proportional $t^{1/2}$) then local reptation takes over (proportional $t^{1/4}$) there after reptation prevails ($t^{1/2}$) and finally translational chain diffusion occurs (proportional t).*

The second process where the chain performs Rouse motion along the tube is called local reptation while the creep like diffusion along the tube, that eventually leads to a complete tube renewal is also termed pure reptation. The terminal time τ_d after which the chain has left its original tube determines to a large extend the viscosity of the melt ($\tau_d \approx \tau_\eta$ see Eq.[29]). Beyond that time reptation diffusion prevails.

As we have alluded to in context with the Rouse Model, in Gaussian approximation the self - correlation function directly relates to the mean square displacement. If the Gaussian approximation would be valid, we would assume that the self-correlation function could be directly interpreted in terms of the m.s.d. However, as Fatkullin and Kimmich [18] have shown, the real process has to be modelled by projecting the segment probability distribution due to the Rouse motion along the random walk like contour path of the tube, this leads to a non Gaussian probability distribution of the segments at times $t > \tau_e$.

$$S_{self}(Q, t > \tau_e) = \exp\left[\frac{Q^4 d^2}{72}\frac{\langle r^2(t)\rangle}{3}\right] erfc\left[\frac{Q^2 d}{6\sqrt{2}}\sqrt{\frac{\langle r^2(t)\rangle}{3}}\right] \tag{32}$$

Thus, the Gaussian approximation for times longer than τ_e is invalid. The effect on the scattering function is, if it is wrongly interpreted in terms of the Gaussian approximation, that the cross over to local reptation appears to occur at a significantly shorter time than τ_e. However, the generic asymptotic behaviour remains untouched. In the sense of Eq. [31] the mean square displacement of a chain segment may still be directly observed by incoherent quasielastic neutron scattering. In the local reptation regime, we expect to observe the predicted cross over from a $t^{1/2}$ to a $t^{1/4}$ law.

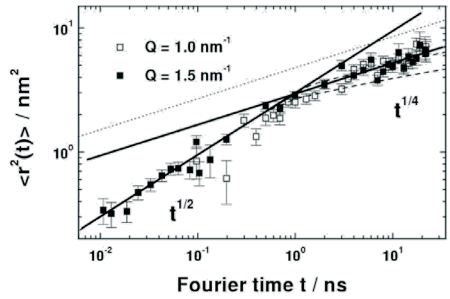


Fig. 16: *Mean square displacements taken from a measurement of the incoherent structure factor from a fully protonated PE Melt assuming the Gaussian approximation. The solid lines describe the asymptotic power laws {$r^2(t)$} proportional $t^{1/2}$, $t^{1/4}$. Dotted lines: predictions from the Gaussian approximation; dashed lines (see text).*

Fig. 16 presents experiments performed on polyethylene samples at a temperature, where also the single chain dynamics structure factor was studied [19]. In Fig. 16 the data are displayed in terms of an effective meansquare displacement, assuming the Gaussian approximation. We nicely observe the predicted cross over from $t^{1/2}$ to a $t^{1/4}$. However, if we insert the dynamic parameters for polyethylene into Eq. [31] we would expect a mean square displacement along

the tube according to the dotted line in Fig. 16. The discrepancy explains itself in considering the non Gaussian character of the curve linear motion according to Eq. [32] (upper dashed line for Q=0,1Å^{-1}

6.2 Single chain dynamic structure factor

We now turn to the conceptionally more demanding question of the coherent scattering from a single labelled chain that is given, by the single chain dynamic structure factor $S_{chain}(Q,t)$. This quantity is strongly affected by topological tube constraints.

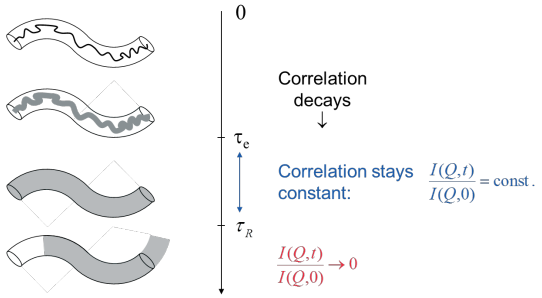


Fig. 17: *Schematic presentation of the various stages in the time development of the single chain dynamic structure factor. At short times unrestricted Rouse Dynamics time takes plac;e beyond $\tau_e$fluctuations fill the tube; For times larger than τ_R the chain creeps out of the tube.*

Fig. 17 visualises the concept, that we will now go through as a function of time.

(i) At short times $t<\tau_e$ the chain will perform unrestricted Rouse motion and the dynamic structure factor for Rouse motion (Equ.[12]) should well describe the dynamics. At short times the tube constrains are not yet effective. In this way the chain explores the lateral constraints set by the tube. Density fluctuations of the chain are laterally equilibrated across the tube profile.

(ii) Once, this is achieved further density fluctuations of the labelled chain will only be possible via Rouse relaxation along the tube. Under such circumstances the structure factor to a first approximation mirrors the form factor of the tube. The correlations will stay and the scattering experiments will reveal the size of the topological constraints.

(iii) In the creep regime $t>\tau_R$ the memory of the tube confinement will be gradually lost and the dynamic structure factor should reveal the fraction of the still confined polymer segments.

(iv) Finally in the diffusive regime the chain reptation diffusion coefficient will be measured.

De Gennes [20] and Doi and Edwards [4] have formulated tractable analytic expressions for the dynamic structure factor. Thereby, they neglected the initial Rouse regime i.e. the derived expression is valid only for $t > \tau_e$ once confinement effects become important. The dynamic

structure factor is composed from two contributions S^{loc} and S^{esc} reflecting local reptation and escape processes from the tube.

$$\frac{S_{chain}(Q,t)}{S_{chain}(Q)} = \left[1-\exp\left(-\frac{Q^2d^2}{36}\right)\right] S^{loc} + \exp\left(-\frac{Q^2d^2}{36}\right) S^{esc} \tag{33}$$

The local reptation part was calculated as

$$S^{loc}(Q,t) = \exp\left(\frac{t}{\tau_0}\right) erfc\left(\sqrt{\frac{t}{\tau_0}}\right) \tag{34}$$

with $\tau_0 = \frac{36}{Wl^4Q^4}$.

A general expression for $S_{esc}(Q,t)$ due to pure reptation was given by Doi and Edwards [4]. For short times $S_{chain}(Q,t)$ decays mainly due to local reptation (first term) while for longer times (and low Q) the second term resulting from the creep motion is important. The ratio of the two relevant time scales τ_0 and τ_d is proportional to N^3. Therefore, for long chains at intermediate times a pronounced plateau in $S_{chain}(Q,t)$ is predicted. Such a plateau is a generic signature for confined motion.

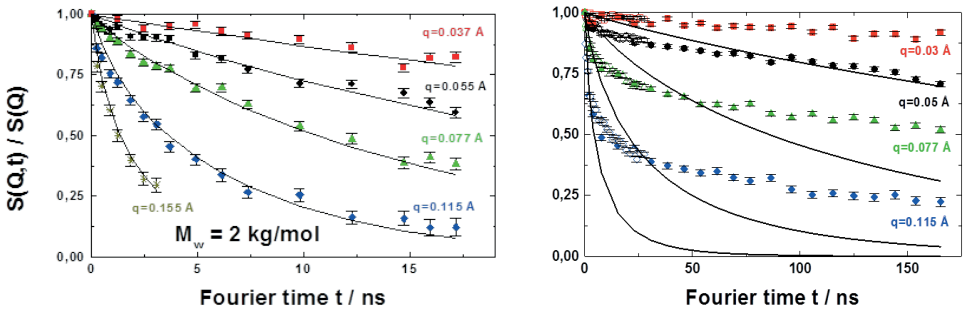


Fig. 18: *Single chain dynamic structure factors from PE melts at 509K (a) M_w=2000; (b) M_w=12400). The solid lines display the predictions of the Rouse Model.*

In order to illustrate the effect of topological constraints of the tube, Fig. 18 compares experimental results from two different polyethylene (PE) melts both studied at 509K.
Fig. 18a displays the results for the short chain melt (M_w=2kg/mol) The solid lines show the prediction of the Rouse dynamic structure factor (Eq. [12])- very good agreement is achieved. Fig. 18b presents equivalent results from the higher M_w melt (M_w=12.4kg/mol). The solid lines again show the predictions of the Rouse model. Also note, that the time scale for the higher molecular weight sample is extended by one order of magnitude. For the long chains the Rouse model fails completely. Only in the short time regime the initial decay of the structure factor is depicted, while for longer times the relaxation behaviour is strongly retarded signifying the tube effects.

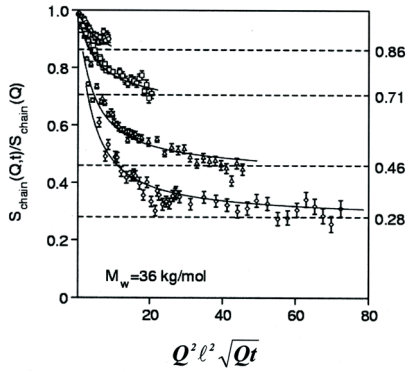


Fig. 19: *Single chain dynamic structure factor from a long polyethylene melt scaled with a Rouse variable. The dashed line described the Debye Waller factor approximation for the long time plateaus (see text).*

Fig. 19 presents the dynamic structure factor from M_w=36kg/mol PE melt. Now as a function of the Rouse variable $(\Omega_R * t)^{1/2}$ [21] other than in Fig. 8, where the scaled data followed a common master curve, here, the data split into different branches that only at a small value of the scaling variable are coming close together. The fact that the data do not follow the common Rouse scaling is the direct consequence of the dynamics length scale of the tube, that invalidates the Rouse scaling properties.
We note, that this length is of dynamical character and can not be observed in static equilibrium experiments. The heights of the achieved plateaus allow a first estimate for the amount of confinement. If we identify the plateau levels with a Debye Waller factor as a measure for the confinement, we get d=46Å, a value, that is a lower estimate for the tube diameter, since S^{loc} is not fully relaxed. The horizontal lines Fig. 19 are the predictions from this Debye Waller factor estimate. A description with the tube dynamic structure factor (Equ.[33]) yields only the slightly higher value of d=48Å.

7 Limits of the Reptation Model

If we compare the predictions of simple reptation for the molecular weight dependences of viscosity and diffusion coefficient with the experimental findings displayed in the chapter on the macroscopic dynamics we realise, that the predictions are only qualitatively in agreement with theory:

(i) Experimentally the viscosity is found to generally follow a power law proportional to $M^{3.4}$ instead of M^3.

(ii) The translational diffusion coefficient is found to behave as $M^{-2.3}$ instead of M^{-2} required by reptation.

(iii) Also the detailed frequency dependence of the dynamic loss modulus G'' does not follow the predicted $\omega^{-\frac{1}{2}}$ behaviour but rather is found to display a $\omega^{-\frac{1}{4}}$ law.

In order to cure the short comings a number of additional relaxation processes were introduced, that are consistent with reptation. The most prominent among them are contour length fluctuations (CLF) and constraint release (CR).

7.1 Contour Length Fluctuations

The CLF effect evolves from the participation of the chain ends in the local reptation process (Fig. 20).

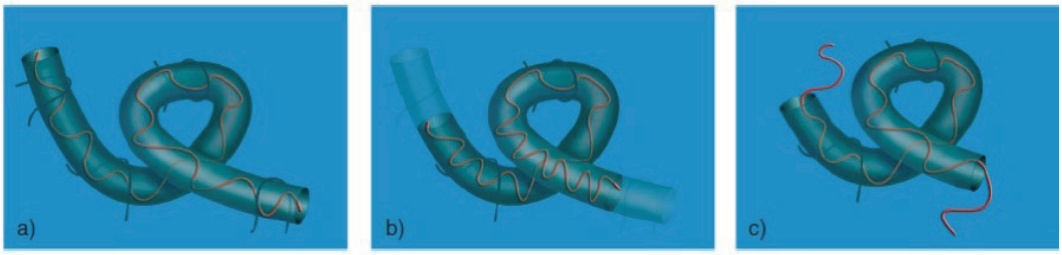


Fig. 20: *Principles of contour length fluctuations*

a) The surrounding chains confine a given chain inside a virtual tube of diameter d
b) The chain ends perform fluctuations retracting into the original tube; thus, forgetting the initial confinement of the vacated tube parts
c) The chain ends may then explore new regions in the melt

Any chain retraction and subsequent expansion will lead to a loss of memory of the original tube confinement. Thus, the tube becomes shorter with time. Mathematically the problem is treated as a first passage problem. Whenever a tube contour s is visited by a free end, the tube ceases to exist. The functional form of the tube survival probability μ(t) was derived theoretically as [22]

$$\mu(t) = 1 - \frac{1.5}{Z}\left(\frac{t}{\tau_0}\right)^{\frac{1}{4}} \tag{35}$$

where Z is the number of entanglements along the chain. Eq. [35] provides quantitative knowledge on the chain fraction that at a time t is still confined. For polyethylene all parameters are known from NSE experiments on asymptotically long chains where the CLF effect does not play a role.

With this knowledge an experiment was designed [23] where the dynamic structure factor of a chain that is subject to CLF was compared with that of an identical chain where the contrast of those segments which within the experimental time frame are affected by CLF was matched.

The experimental idea is displayed in Fig. 21: case (a) is realised by performing an experiment on a fully protonated chain in a deuterated matrix. In this case, the full chain dynamics including the CLF is observed. Case (b) is realised by a chain, where the inner part is protonated and the two outer chain sections of a length, that is affected by CLF are

deuterated and therefore not visible in the deuterated matrix. In such a case the dynamics of the chain should be equal to that of an asymptotically long fully confined chain.

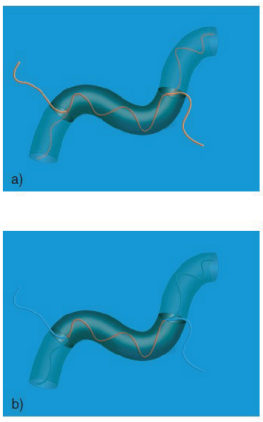


Fig. 21:

a) For the fully protonated chain in a deuterated matrix local reptation and contour lengths fluctuations are visible

b) For a center labelled (protonated chain with deuterated chain ends) in a deuterated matrix only local reptation is visible. The contour lengths fluctuations of the chain ends are masked out.

With the known parameters for PE Eq. [35] reveals that on average on each side 220 monomers are released during the observation time of 190 ns at 509 K. The experiments were performed on two different chains of a molecular weight of 25kg/mol one was fully hydrogenated and the other had deuterated labels of about M_w=4kg/mol corresponding to 260 monomers on each end. Both were studied in a deuterated matrix of the same molecular weight.

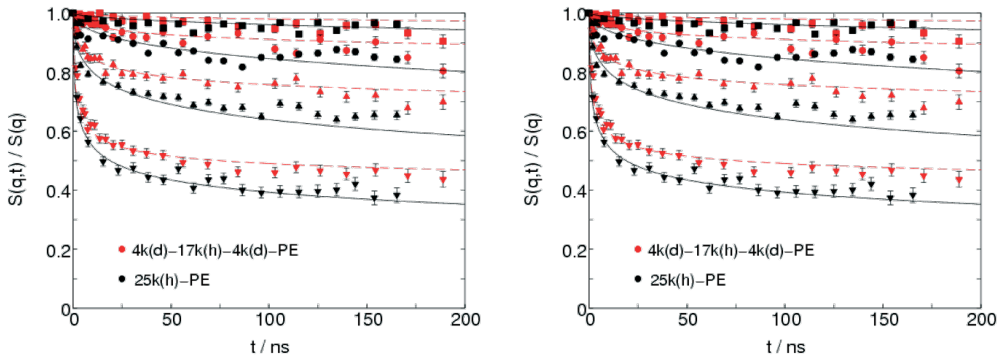


Fig. 22: *(a) Comparison of the single chain dynamic structure factors from an end labelled chain and a fully labelled chain. The solid and dashed lines represents fit with the corresponding models.*

(b) Comparison of the single chain dynamic structure factors for a very long and an end labelled polyethylene chain. The inner part of the shorter end labelled chain relaxes as the very long chain.

Fig. 22 presents the normalized dynamic structure factors S(Q,t)/S(Q) as a function of time for different Q values. Fig. 22a compares the results for the differently labelled chains. The fully labelled chain relaxes significantly stronger than the corresponding center labelled counter part. Apparently the constraints are stronger for the center labelled chain than for the chain where the outer parts are also visible. We also note, that in the case where the ends were masked the chain center part shows exactly the same structure factor as that for a very long chain where CLF processes do not play a role (Fig. 22b). This result directly demonstrates that the action of CLF happens at the chain ends and that the confinement remains fully conserved with time in the center. The agreements of the data sets in Fig. 22b also show, that constraint release in the investigated space time regime does not play any role. The solid lines in the two figures represents theoretical descriptions of the CLF effect and are taken from reference [22].

At the end we like to emphasize, that CLF also affects the macroscopic melt properties in a significant way. (i) It has been shown, that CLF introduce the $\omega^{-\frac{1}{4}}$ regime into the spectrum of G''(ω). However, the other limiting mechanism (CR) adds further significant modifications. (ii) The translational chain diffusion necessarily is also affected. First, the terminal time defining the diffusion step is reduced since reptation has only to relax the not yet released central parts of the tube and secondly the diffusive lengths is reduced, since only the displacement of the central part counts. Both effects do not cancel and the net effect gives D proportional $(N/N_e)^{-2.4}$ in very good agreement with experiments. (iii) Finally also the anomalous power law exponent of 3.4 for the viscosity mass relation has been attributed to CLF.

7.2 Constraint Release

As we have discussed above, contour length fluctuations are an effect of the confined chain itself. On the other hand constraint release (CR) stems from the movement of the other chains building the tube that of course undergo the same dynamical processes as the confined chains. This is an intrinsic many body phenomena and much more difficult to treat than CLF. Topological constraints for a polymer can be released (or created) by the reptation of surrounding polymers as shown in Fig. 23.

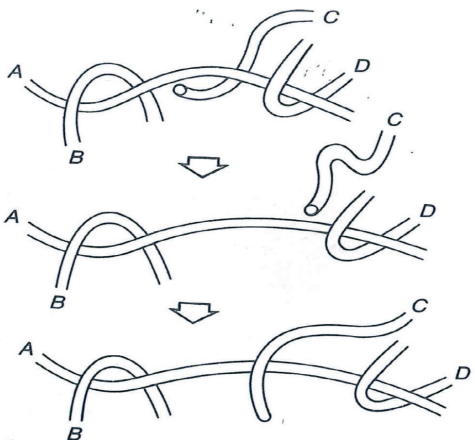


Fig. 23: *Release and creation of the topological constraints. The topological constraints imposed on the chain A by C is released and recreated by the motion of C [4].*

Such a process will cause a conformational change of the tube at the point where a chain has moved in or out. The simplest possibility to describe this process is to regard the conformational tube change as a local jump of the tube. Since the jump rate would then be in the order of $1/\tau_d$, this process has a negligible effect on the longest relaxation time τ_d itself – the longest relaxation time of the Rouse tube dynamics would be $\tau_d(N/N_e)^2$ for beyond the life time of the tube. However, the process provides an additional relaxation to the dynamic modulus in the plateau regime. And this is important for the quantitative interpretation of rheological results.

In order to access this process on the chain level we need to study the dynamics of a probe chain in matrices of different molecular weights. The smaller the molecular weight of the matrix, the more often constraint release processes will take place and will change the dynamic structure factor of the probe chain [24].

In order to separate the effect of CR from the CLF process, the probe chain has to be long enough such that end effects like CLF do not play a role in the experimental window that is accessible. Furthermore the concentration of the probe chains needs to be low enough that interaction effects in between probe chains are negligible. Experimentally this situation was realized in considering polyethylene probe chain of M_w = 36 kg/mol that were successively mixed with matrix chains in between 36 and 1 kg/mol – the entanglement molecular weight for PE is M_e = 1-2 kg/mol.

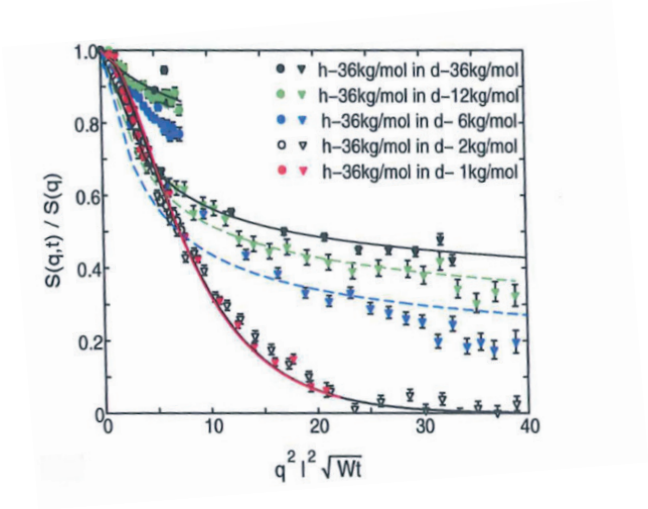


Fig. 24: *Dynamic structure factor of a long labeled chain (M_w = 36 kg/mol) in different shorter matrix chains (M_w = 36, 12, 6, 2, 1 kg/mol as indicated in the plot) in a Rouse scaling representation for two different Q values (circles 0.5 nm^{-1}, triangles 1.15 nm-1).*

Fig. 24 displays the experimental results for the 36 kg/mol chain in matrices of different molecular weight. The data display a clear transition from confined reptation like motion in higher molecular weight matrices to free Rouse motion in short matrix chains. This is demonstrated in the Rouse scaling representation, that is sensitive to the occurrence of the reptation confinement length scale. For the long chain in a matrix of the same molecular

weight a confinement as for infinite long chains is observed. Thus, for a 36 kg/mol chain neither CLF nor CR play a significant role in the accessible NSE time window. The chain remains confined inside the tube (black points in Fig. 24).

With decreasing matrix length (12 kg/mol, 6 kg/mol) an increasing loss of confinement becomes visible which shows itself in a stronger decay of the dynamic structure factor compared to the high molecular weight matrix. The additional relaxation reflects the phenomenon of constraint release: The loosening of the tube confinement due to the motion of the surrounding chains. In the scaling representation the splitting of different Q-values is still evident but the reptation model fails to describe the data. For the 12 kg/mol matrix a fit with the reptation model results in a wrong tube diameter for the 6kg/mol matrix the model also fails qualitatively (dashed line).

Eventually for short chain matrices of only one entanglement length (about 2 kg/mol) and below (1kg/mol) the long labeled chain displays Rouse scaling. Obviously the matrix chains are too short to confine the long chains.

The key question now is which dynamical processes underlie the CR process and lead to the gradual loss of confinement with decreasing matrix chain length. The reptation time of the matrix chains (τ_d(12 kg/mol) ≈ 5000 ns, τ_d(6 kg/mol) ≈ 500 ns) is far beyond the experimental time range. This holds even more for the characteristic times for constraint release $\tau_{CR} \approx \tau_e \left(M_{long} / M_e \right)^2 \left(M_{matrix} / M_e \right)^3$. They are in the range of hundreds of microseconds. But even for the 12 kg/mol matrix which is well entangled the effect of CR is considerable.

This fact demonstrates that for an estimation of the characteristic time scale all dynamical processes that are known to determine the segmental motion have to be taken into account. These processes include next to the reptational creep also CLF of the chain ends. In contrast to the center of mass, the segmental meansquare displacement is significant in the experimental time range. Using the experimental parameters it can be estimated within the Rouse model to $\left\langle r_{segm}^2 \right\rangle = \sqrt{4/\pi Wt^4 t}$. After the initial decay of S(Q,t=20-30ns) the segmental displacement becomes about 4.0 nm.

In order to clarify this further, an experiment was designed in order to separate the different dynamic processes. For a long chain in a short matrix CLF is negligible. Therefore the effect of CR of the matrix chain can be observed. On the other hand for the dynamics of a short chain in a long matrix, the CLF of the short chain are dominating. No CR of the matrix chains can occur. Comparing the dynamic structure factor of such a corresponding pair of samples, the contribution from CLF and CR of the shorter chains can be separated.

Fig. 25 shows the result for a 12 kg/mol chain in a 36 kg/mol matrix and vice versa. The long chain relaxation by constraint release of the short matrix is obviously identical to the relaxation of the confined short chains by contour length fluctuation. Thus, in the 12 kg/mol matrix the CR visible in the long chain dynamics may be traced solely to the CLF of the matrix chains. Thus, this experiment shows that even CLF alone can cause CR effects.

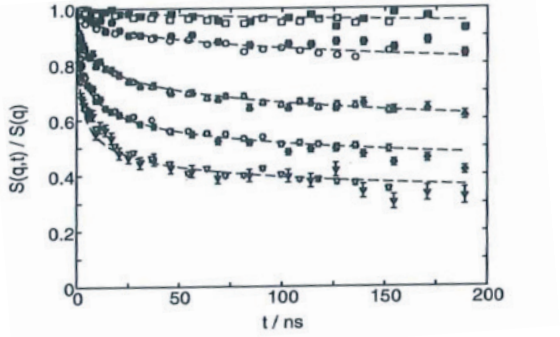


Fig. 25: *Dynamic structure factor of a 36 kg/mol chain in a 12 kg/mol matrix (solid symbols) and vice versa (open symbols). Q values (in nm^{-1}): squares 0.3, circles 0.5, up-pointing triangles 0.77, diamonds 0.96, and down-pointing triangles 1.15. Lines are just guides for the eye.*

8 Dynamics of Ring Polymers

Within the tube confinement that is build by surrounding chains, linear chains relax by leaving their initial tube. Branched polymers relax by backfolding the branches against entropic barriers. In all these processes the chain ends play an essential role. In this context ring polymers are of particular interest since they exhibit a unique topology: rings are closed structures without ends. In dense melts this feature significantly influences both ring conformations and dynamics. Interpenetration of ring polymers is entropically hindered and therefore rings are predicted to prefer the conformations of a crumbled globule or that of a lattice animal. The lack of ends changes the dynamics qualitatively. All reptation related processes become impossible and qualitatively different motional mechanisms are expected. This explains the intense scientific interest in the dynamics of ring polymers.

Since the synthesis of well-defined and pure large ring polymers is highly demanding, recently there has been a focus on simulations of their structure and dynamics. MD studies on bead and spring rings concluded a rather compact structure and an asymptotic conformation of a crumbled globule for large rings. The studies unraveled a subdiffusive center of mass (c.m.) behavior $\left\langle r_{cm}^2(t)\right\rangle \sim t^{3/4}$ at early times, before the transition to normal translational diffusion takes place at about $\left\langle r_{cm}^2\right\rangle = 2.5R_g^2$. At higher molecular weights D ~ N^{-2} was found. Regarding the segmental relaxation, simulations observe a significant slowing down towards $\left\langle r_{seg}^2(t)\right\rangle \sim t^{\nu}$ with ν = 0.25 – 0.35. Such a power law is also characteristic for the local reptation regime where the Rouse modes relax within the stiff confining tube yielding $\left\langle r_{seg}^2(t)\right\rangle \sim t^{1/4}$. The reason for the occurrence of such a time regime for rings is unclear.

Recently the synthesis of well-defined large polyethylene oxide (PEO) rings became possible yielding rings in a molecular weight range from 2 kg/mol to 20 kg/mol [25].

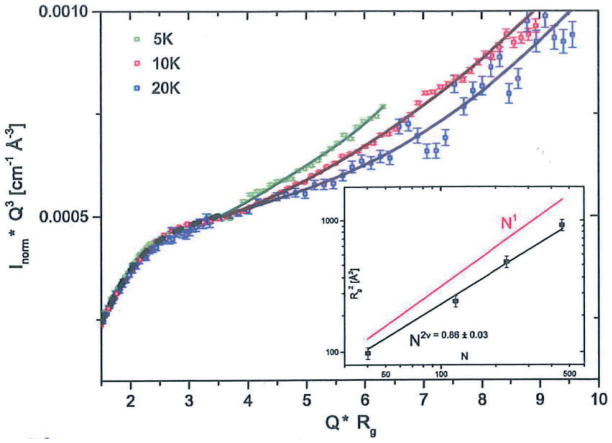


Fig. 26: *SANS data for the 5 K (green symbols), 10 K (red), and 20 K (blue) rings plotted versus QR_g. The data yield a perfect master curve in the low QR_g regime. The dashed black line up to QR_g = 3.5 represents a fit to the data, the colored lines in the high QR_g regime are guides for the eyes. The inset shows the N dependence of R_g^2 (red line: Gaussian expectation, black line: experimental data including one additional small 2 K PEO ring.*

Fig. 26 shows small angle neutron scattering data for 5 K, 10 K and 20 K rings in a representation S(Q)*Q^3 versus QR_g. In the low QR_g regime all data fall perfectly on one master curve if properly normalized for volume fraction and number of monomers N. As predicted by the crumbled globule picture and simulation in the intermediate Q-range a shoulder develops indicating an approach towards an asymptotic mass fractal behavior $R_g \sim N^{1/3}$. With increasing N the shoulder becomes more pronounced and extended indicating the increasing local compactness. This result is substantiated by the N dependence of R_g that is plotted in the lower inset of Fig. 26. For all N the conformation is significantly more compact than the Gaussian prediction (red line). For intermediate Q a power law behavior in the form factor following $Q^{-1/\nu}$ with ν = 0.43 ± 0.015 was found. With this Flory index the mass dependence of R_g^2 is very well described (see black line in Fig. 26).

Fig. 27 presents the dynamic structure factors obtained by NSE for the 10 and 20 K rings. These data reveal both, the internal dynamics and the c. m. motion up to time scales where $\langle r_{cm}^2 \rangle$ well exceeds R_g^2. The selection rules for internal modes for ring polymers (only even modes) allow a direct extraction of the center of mass mean square displacements at low Q. There, even for the large ring internal modes do not contribute to S(Q,t). Thus, NSE allows direct measurements of the center of mass displacement via $S(Q,t) = S(Q)\exp\left(-Q^2 \langle r_{cm}^2(t) \rangle / 6\right)$. The validity of this Gaussian assumption is proven by the Q dependence of the low Q spectra.

The long time translational ring diffusion can be measured by pulse field gradient NMR that reveals the translational diffusion coefficient in the millisecond regime. In Fig. 28 all center of mass m.s.d scaled with N are presented as a function of time. At times just below those

associated with the normal diffusion regime the scaled data collapse to a single master curve displaying a subdiffusive center of mass motion with a m.s.d. proportional to $t^{3/4}$. The mass scaling seems to indicate purely friction controlled subdiffusive motion. The crossover from subdiffusive to normal diffusion occurs at about $\tau_{cross} \approx 2.5 R_g^2 / D$ while for the small 5 K ring the crossover is shifted towards $6R_g^2$.

While the evaluation of the center of mass m.s.d. is straight forward, the access to the segmental motion i.e. the internal dynamics is much more involved and challenging. In this lecture we will not go through the different steps of data evaluation but report the major results.

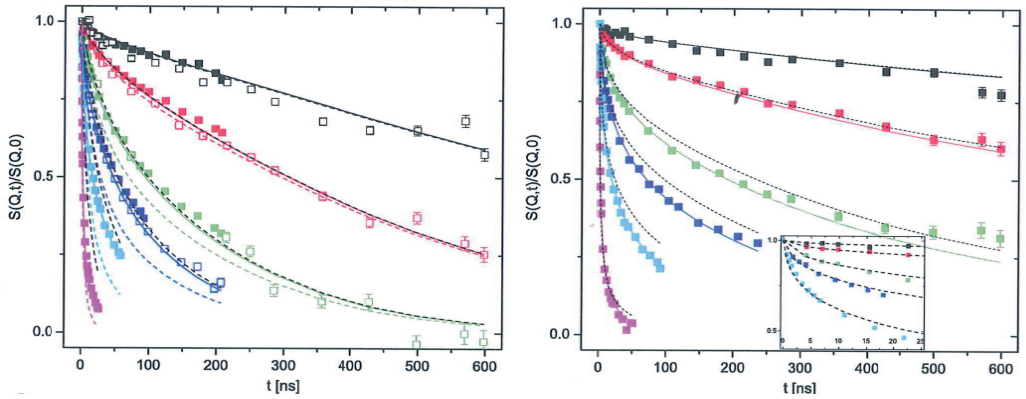


Fig. 27: *NSE spectra for the 10 K (left panel) and 20 K (right panel) ring for Q values (from top to down) 0.03, 0.05, 0.08, 0.1, 0.13, and 0.2 Å^{-1} at T = 413 K. Full and empty symbols (left panel) refer to two different wavelength setups. Lines and inset, see text.*

In Fig. 27 different lines represent the different contributions of motion to the dynamic structure factors. The dashed black lines present the contribution of center of mass motion. We see that for the 10 K Ring center of mass motion describes very well the first three low Q results, thereby emphasizing the Gaussian character of the diffusion process. For the 20 K ring diffusion alone describes the data at 0.03 and 0.05 Å^{-1}, while at larger Q significant deviations towards faster decays are evident. These faster relaxations relates to the internal motion.

Now one could assume that the internal dynamics of the rings is described by the Rouse dynamic structure factor modified for rings [26].

$$S(Q,t) = \frac{1}{N}\sum_{i,j}\exp\left[-\frac{\langle r_{cm}^2\rangle Q^2}{6} - \frac{1}{6}Q^2 l^2 |i-j|^{2\nu}\left(1-\frac{|i-j|}{N}\right)^{2\nu} - \frac{4Q^2N^{2\nu}l^2}{6\pi^2}\sum_{P\min,even}^{N}\frac{1}{p^2}\cos\left(p\pi\frac{(i-j)}{N}\right)\left(1-\exp\frac{-tp^2}{\tau_R}\right) - F\frac{4Q^2N^{2\nu}l^2}{6\pi^2}\sum_{p,even}^{p_{\min}-2}\frac{1}{p^2}\cos\left(p\pi\frac{(i-j)}{N}\right)\left(1-\exp\frac{-tp^4}{\tau_{cross}}\right)\right] \quad (36)$$

The sum in Equ.[36] over the indices i, j runs over all monomer coordinates of the ring, the sum over p in the exponent addresses all Rouse modes (Equ.[9]) – for a ring only even modes p contributes. The second sum will be addressed later.

For p_{min}=2 the last term vanishes and equation [36] yields the dynamic structure factor for Gaussian rings. Describing the internal ring dynamics in this way for the 10 K ring the colored dashed lines result. It is clear that the internal dynamics is strongly overestimated. Thus not all Rouse modes are contributing to the internal relaxation. In order to assess to what extend Rouse relaxation takes place, the data were analyzed in terms of a reduced number of Rouse modes. Suppressing the first ring mode p = 2 and activating all other modes with $p \geq 4$ yields an excellent description of the 10 K ring data. (Solid lines in Fig. 27 left) In this fit only p_{min} was a free parameter. From the wavelength of the first active mode we conclude that ring sections of about 60 monomers N / p_{min} relax undisturbed by topological effect. For the 20 K ring the initial decay of the structure factor (see insert in Fig. 27 right) may be described by a similar procedure. In this case p_{min} = 8 results. It is remarkable that this p_{min} again refers to a ring section of about 60 monomers that can relax without topological hindrance. However, if we look on the long time behavior, then this is not the full story. There are significant deviations from the model at times beyond 20 ns. A way out would be to allow longer wavelength modes but this would significantly overpredict the decay at short times. Thus, we conclude that a different further relaxation mechanism takes place, however, under significant constraints.

We may interpret the result for the early motion as a free Rouse relaxation of loops formed e. g. in a lattice animal structure. These loops would have a molecular weight of about 2.5 kg/mol and are able to relax freely. Incidentally this loop size is comparable with the entanglement length for PEO of M_e = 2 kg/mol. While for the 10 K ring the picture of loop relaxation is sufficient to describe the observed spectra, this is clearly not the case for the 20 K ring. Here loop relaxation alone underestimates the internal relaxations of the ring.

In order to go further, we use the result of a recent simulation revealing a $\sim t^{-0.3}$ evolution of the ring segmental dynamics until the crossover to normal translational diffusion is reached. In terms of a mode picture in a phenomenological way such a $t^{1/4}$ behavior of the m.s.d. is achieved in adopting the power law for the characteristic times $\tau_p \sim p^{-2}$ to $\tau_p \sim p^{-4}$. Generally and asymptotically a p^{-n} power law for the mode relaxation leads to a m.s.d. proportional to $t^{1/n}$. Following this idea for the lower mode numbers, the dispersion of the characteristic times τ_p should be changed to p^{-4}.

While for the fast Rouse relaxation the basic time is given by the Rouse time τ_R, at this point a new basic time $\tau_{ring} \approx \tau_{cross}$ would come into play as a reference for the slower process. This time is known from the transition of the center of mass displacements to normal diffusion (Fig. 28).

Fixing τ_{cross} = 1725 ns and fitting the range of spatial relaxation that is achievable for the internal motions of the ring, a very good fit of the 20 K ring data (see Fig. 27 right panel) is achieved predicting a maximum extend of the internal relaxation of about 90 % of R_g.

In Fig. 29 we display the time dependent meansquare displacements of the 20 K ring as it follows from the description of the dynamic structure factor. The blue line displays the time dependent loop relaxation, the red line displays the dynamics of the loops derived from the

motions of the slowly decaying low p modes. At short times this process is marginal but at intermediate times it takes over and dominates the weaker power law for motion of the loops along the lattice animal. Finally, around τ_{cross} it saturates – from here on translational diffusion of the whole ring takes over. The sum of the two processes in Fig. 29 yields a slope of 0.32 at intermediate times in almost perfect agreement with simulation (slope 0.35).

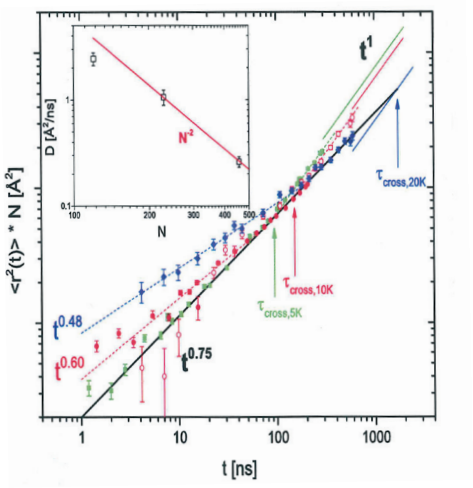


Fig. 28: *Center of mass mean squared displacement. N versus time for low Q = 0.05 Å^{-1} for the 5 K (green), 10 K (red) and 20 K (blue) rings. The inset shows the molecular weight dependence of the diffusion coefficient approaching D α N^{-2} for high N.*

To conclude (i) the ring conformation displays a tendency towards mass fractal behavior that evolves systematically with increasing molecular weight, (ii) a subdiffusive center of mass motion at sub molecular distances with a m.s.d. $\sim t^{3/4}$ is established, (iii) at short times the internal dynamics is dominated by relaxing ring sections or loops that move like short unentangled chains and (iv) at later times slow loop motions following a $\sim t^{\nu}$ law with ν = 0.32 is found.

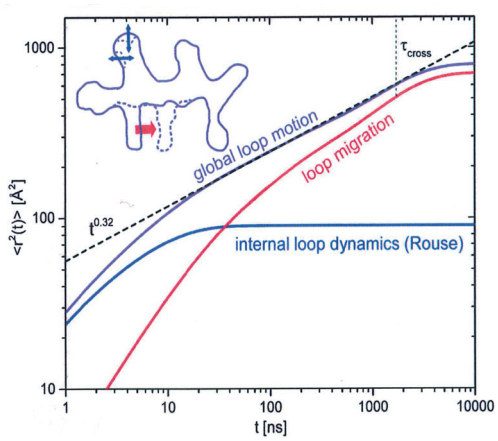


Fig. 29: *Time dependent m.s.d. of the 20 K ring. Blue line: loop relaxation at short times. Red line: dynamics of the loops derived from the motions of the slowly decaying low p modes. Purple line: sum of both with a slope of t$^{0.32}$ at intermediate times. This slope is indicated by the dashed black line.*

9 The Effects of Surfaces and Confinement

The interest in the investigation of polymers under nanoconfinement has been amplified recently by the rising of nanotechnology that aims to create new properties in modifying materials at the nanoscale. Polymers are of particular interest since they offer a large range of applications such as coatings, lubrication, nanocomposites etc. Close to a confining surface the conformations of a polymer are significantly restricted. In addition the interactions with the surface will strongly affect the dynamics. Experimental results on polymers close to surfaces have been interpreted in terms of the formation of a glassy polymer layer close to the surface. Furthermore, the existence of in interface with properties between those of the glassy layer and a bulk has been hypothesized. Looking in particular on nanoparticles dispersed in a polymer matrix it was found that the addition of nanoparticles that interact with a polymer matrix induce dramatic property changes for the resulting polymer nanocomposite. Theoretical work and computer simulations of chain adsorption as the function of adsorption strength reveal the existence of different chain conformations including trains, loops and tails.

In this chapter we will discuss an investigation on the dynamics of polydimethylsiloxane (PDMS chains) confined in anodic aluminum oxide (AAO nanopores)[27]. Fig. 30 displays the system. Fig. 31a and b show an electron micrograph of the AAO nanopore system. We see a hexagonally arranged pattern of nanopores that extend by more than 100 μm into the third dimension. The lower part of the figure presents the polydimethylsiloxane chain; indicated are possible hydrogen bonds with the OH-groups at the AAO surface.

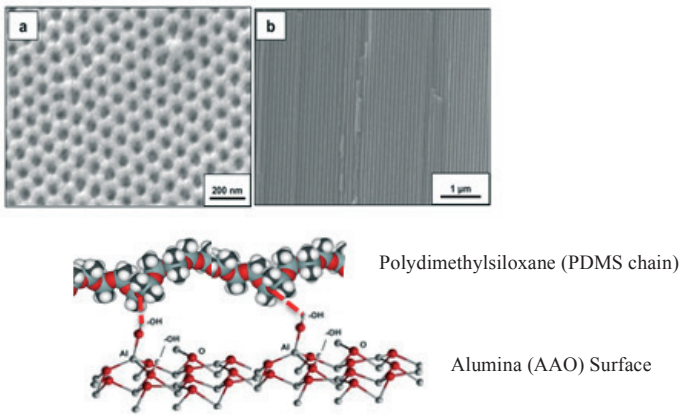


Fig. 30: *(a,b) Electron micrograph of anionic aluminium oxid (AAO) nanopores, (c) the interaction of a PDMS chain with the AA= surface OH groups via H-bonds.*

PDMS has an entanglement molecular weight of M_e =12 kg/mol leading to an entanglement spacing of about 8 nm. The polymers under investigation had a molecular weight of 17.4 kg/mol and were basically non-entangled.

Fig. 31 displays characteristic NSE spectra taken at different Q values over a time range of about 150 ns. The data are characterized by an initial fast decay that is succeeded by a much weaker decay at longer times.

The evolution of a plateau in the single chain dynamic structure factor indicates the presence of a non-decaying part in the correlation function and is a signature of confinement similar to what has been seen for chains confined in a tube in the reptation picture. If we would take the presence of a non-decaying part in the pair correlation function as an indication of an immobilized polymer layer in the vicinity of the pore surface, then these relative contributions to the structure factor should be constant with Q. Thus, already at a very first glance the data disprove the existence of a glassy surface layer that was often invoked e.g. in the interpretation of the rubber elasticity of filled rubbers in terms of a so called bound layer.

After this first insight we now quantify the observations. In a first approximation we describe the long time tales in $S_{chain}(Q,t)$ by a Q-dependent plateau α(Q). Thus, we have

$$S(Q,t)_{chain} = \left[1-\alpha(Q)\right] S_{Rouse}(Q,t) + \alpha(Q) \tag{37}$$

where S_{Rouse} (Q,t) is the dynamic structure factor originating from the Rouse model (see Equ.[12]).

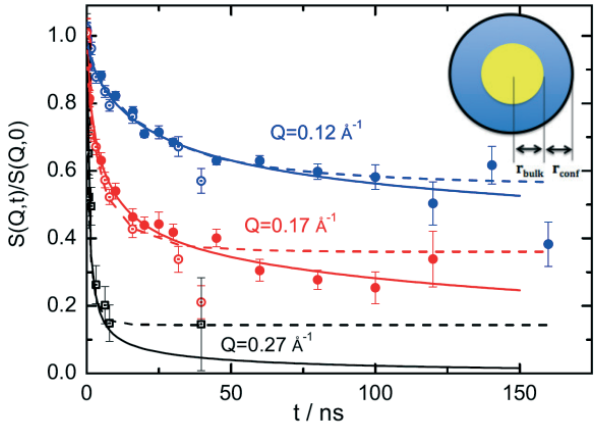


Fig. 31: *NSE results for confined PDMS. Dashed lines show the fitting of the data using Equs. [37] and [39]. Solid lines present the fitting using a continuous transition region from suppressed to free Rouse models (see text). A sketch of the two-phase model is shown in the inset.*

With this approach we arrive at the dashed lines in Fig. 31. As already seen qualitatively the plateau α(Q) strongly depends on the momentum transfer Q. In a reptation picture the presence of a Q dependent plateau in the dynamic structure factor of a polymer melt is associated with an effective confinement length, the tube diameter, in this model the plateaus are related to the tube diameter by

$$\alpha(Q) = \exp\left(-\frac{Q^2 d^2}{36}\right) \tag{38}$$

With that we can estimate a characteristic confinement length d from the plateau heights. Values are found between 3.1 and 3.9 nm, significantly smaller than the tube size in bulk PDMS (approximately 8 nm). In addition the effect that the characteristic confinement length is again Q-dependent indicates that a description of the chain dynamics by means of a reptation like approach is not valid. Thus, the model needs to be sharpened: The model should contain a fraction of free bulk like chains. Many chains are far from the surface and are not expected, to be affected in any way. The second fraction of chains that is close to the surface, however, is assumed, to be effectively confined. The dynamic structure factor is presented by the formula

$$S(Q,t)_{chain} = A_{bulk} S_{bulk}(Q,t) + A_{conf} S_{conf}(Q,t) \tag{39}$$

where A_{bulk} and $A_{conf} = 1 - A_{bulk}$ are the fractions of the bulk and confined phases respectively. $S_{bulk}(Q,t)$ is the dynamic structure factor of the free chain defined by Equ.[12] and $S_{conf}(Q,t)$ is the dynamical structure factor of the confined chain that is modeled by a modified Rouse Ansatz, where only Rouse modes above a minimum value of p contribute (Equ.[12] with $p_{min} > 1$). Thus, the effective confinement is taken into account by suppressing the center of mass diffusion and the first modes for $p < p_{min}$ with wave lengths longer than the distance between chain adsorption points.

Furthermore, the anchoring conditions will be not identical for all chains. Therefore, a smoothly varying cut off function (modeled by a Fermi function) was employed, that increases the mode relaxation rate gradually from virtually 0 to the standard Rouse model rate as a function of p. The uncertainty in p_{min} becomes 1.1 and the resulting p_{min} amounts to about 6. The result of this description is shown as solid lines in Fig. 31. It describes the data very well. The mode $p_{min} \approx 6$ has a wave length corresponding to 38 segments. The end to end distance of a sub chain corresponding to 38 segments gives an effective confinement length $d = \sqrt{\langle R_{E,subchain}^{\,2} \rangle} = \sqrt{N/p_{\min}\, l^2} = 3{,}4\,nm$. Furthermore it turns out that 75 % of all polymers are confined.

The obtained distribution of bulk and confined chains can be transferred to a corresponding layer thickness. We find a layer of confined chains with the thickness $R_{conf} = 6.5$ nm and the cylindrical bulk phase again with a radius of $R_{bulk} = 6.5$ nm (see insert in Fig. 32). The effect that the size of a PDMS sub chain between the encoring points is 3.4 nm clearly indicates that the layer with a thickness of 6.5 nm is not only built by chains adsorbed a the surface but rather support the situation as illustrated in Fig. 32.

The polymer chains anchored to the surface form loops. Neighboring chains then can interpenetrate these loops. These chains then may only diffuse by reptation. Estimating the escape time based on a confinement length of 3.4 nm yields to a reptation time 60 times longer than the maximum observation time. It follows that neither the anchored nor the entangled chains exhibit center of mass diffusion in the time frame of the experiment. These penetrating chains can be considered as the interphase between the polymer adsorbed on the surface and the bulk polymer phase. The interphase forms as a consequence of confinement and dramatically changes the properties compared to those of the unentangled polymer melts.

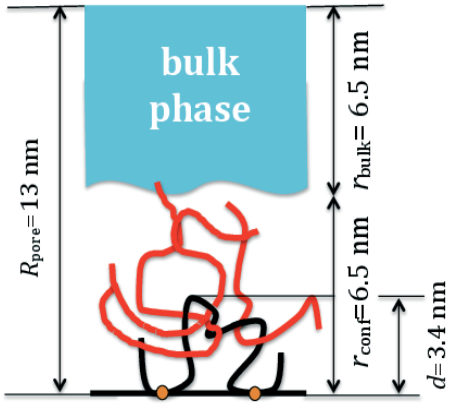


Fig. 32: *Schematic representation of the artificial surface-induced entanglements in the confined polymer melt. The black line represents the chain adsorbed on the surface of an AAO nanopore, and the red lines show entangled chains in the confined phase.*

To conclude the dynamic behavior of PDMS under confinement in nanopores follows a two phase model: One free bulk like fraction of chains and one phase of confined polymers. This phase is characterized by a vanishing center of mass diffusion and by a suppression of long wave length Rouse modes as expected for multiply anchored and topologically confined chains. The corresponding confined layer of 6.5 nm thickness on the one hand consists of a dominant fraction of highly mobile segments and on the other hand is evidence for the presence of a polymer interphase induced by the interaction of the anchored polymer and the surrounding melt.

10 Conclusion and outlook

We have presented some representative results from neutron spin echo spectroscopy on the dynamics of macromolecules. We have displayed recent results on the universal dynamics of flexible polymers from the entropy driven Rouse dynamics to confinement and reptation. The lecture attempted to transmit a flavour of what can be achieved with high resolution neutron spin spectroscopy that permits access to the molecular motion simultaneously in space and time.

After some brief description of generic results on the macroscopic dynamics of polymer melts the lecture commenced with a description of the standard model of polymer motion, the entropy driven dynamics covered by the so called Rouse model. In the spatial range where the Rouse approximations are valid, the NSE measurements have confirmed most of the predictions of the Rouse model both for the self- and pair correlation function. We then presented the changes in the Rouse dynamics that occur in solution, where hydrodynamic interactions are important. Under these circumstances the chain dynamics changes in a qualitative way resulting in a scaling behaviour of the relevant observables that differs from the Rouse model.

Towards larger scales topological interactions resulting from the mutually interpenetrating chains gain dominating influence and confine the chain motion to a tube along the chain profile. We have presented measurements on the dynamic structure factor of a reptating chain which unequivocally confirm the picture of local reptation i.e. Rouse relaxation along the contorted tube. A measurement of the self correlation function corrobates the picture. To conclude this section we have described results on contour length fluctuations and constraint release that modify the reptation mechanism significantly.

Finally, very recent new results on the dynamics of polymer rings and the effect of attractive surfaces on the melt dynamics were presented. The ring dynamics addresses the fundamentals of polymer motion. The unique ring topology prohibits the standard polymer motions, that relies on the existence of chain ends and poses novel challenges. With the aid of NSE experiments a special centre of mass motion relating to the behaviour of lattice animals and an internal dynamics consisting of the free Rouse motion of loops and a slow loop migration could be unravelled. The effect of attractive surfaces on the chain dynamics is an important ingredient of the advantageous properties of composites. The NSE experiments have provided a better molecular understanding of the surface layer properties

11 References

[1] For a recent review see: D. Richter, M. Monkenbusch, A. Arbe, J. Colmenero, Adv. Polymer Sci. 174, Springer, Berlin, Heidelberg, New York (2005)
[2] M. Doi, S.F. Edwards, J. Chem. Soc. Farad. Trans. 274, 1789; 274, 1802; 275, 38 (1978)
[3] P.G. de Gennes, J. Chem. Phys. 55, 572 (1971)
[4] M. Doi, S.F. Edwards*, The theory of polymer dynamics,* Clarendon, Oxford (1986)
[5] D. S. Pearson, L. J. Fetters, W. W. Graessley, G. V. Strate, E. v. Meerwall, Macromolecules, 21, 711 (1994)
[6] T. Lodge, Phys. Rev.Lett., 83, 3218 (1999)
[7] P.R. Rouse, J. Chem Phys. 21, 1272 (1953)
[8] F. Mezei (Ed.), Neutron spin echo lecture notes in physics 128, Springer, Berlin, Heidelberg, New York (2005)
[9] D. Richter, B. Ewen, B. Farago, T. Wagner, Phys. Rev. Lett., 62, 2140 (1989)
[10] H. Montes, M. Monkenbusch, L. Willner, S. Rathgeber, L.J. Fetters, D. Richter, J. Chem. Phys. 110, 10188 (1999)
[11] J.D. Ferry, *Viscoelastic properties of polymers,* John Wiley &Sons, New York (1980)
[12] L.J. Fetters, D.J. Lohse, D. Richter, T.A. Witten, A. Zirkel, Macromolecules 27, 4639 (1994)
[13] B. Ewen, D. Richter, Advances in Polymer Science, 134, 1 (1997)
[14] E. Dubois Violette, P.G. deGennes, Physics(USA) 3, 181 (1967)
[15] P.G. de Gennes , *Scaling Concepts in polymer physics*, Cornell University Press, Ithaca (1979)
[16] T. Csiba, G. Jannink, D. Durand, R. Papoular, A. Lopp, L. Aurray, F. Bowe, J.P. Cotton, R. Borsali, J. Phys. II 1, 381 (1991)
[17] K. Kremer, G. S. Crest, J. Chem. Phys. 92, 5057 (1990)
[18] N. Fatkullin, R. Kimmrich, Phys. Rev. E 59, 3273
[19] A. Wischnewski, M. Monkenbusch, L. Willner, D. Richter, B. Farago, G. Kali, Phys. Rev. Lett. 90, 058302 (2003)
[20] P.G. de Gennes, J. Phys. (Paris) 42, 735 (1981)
[21] P. Schleger, B. Farago, A. Kollmar, C. Lartigue, D. Richter, Phys. Rev. Lett. 81, 124 (1998)
[22] N. Clarke and T. C. B. McLeish, Macromolecules 26 5264 (1993)
[23] A. Wischnewski, M. Monkenbusch, L. Willner, D. Richter, A. E. Likhtman, T. C. B. McLeish and B. Farago, Phys. Rev. Lett. 88, 058301 (2002)
[24] M. Zamponi, A. Wischnewski, M. Monkenbusch, L. Willner, D. Richter, A. E. Likhtman, G. Kali, B. Farago, Phy. Rev. Lett. 96, 238302 (2006)
[25] R. Gooßen, A.R. Bras, M. Krutyeva, M. Sharp, P. Falus, A. Feoktystov, U. Gasser, W. Pyckhout- Hintzen, A. Wischnewski, D. Richter, Phys. Rev. Lett. 113, 168302, (2014)
[26] G. Tsolou, N. Statikis, C. Baig, S. Stepanou, V.G. Marantzas, Macromolecules, 43, 10692 (2010)
[27] M. Krutyeva, A. Wischnewski, M. Monkenbusch, L. Willner, J. Maiz, C. Mijangos, A. Arbe, J. Colmenero, A. Radulescu, O. Holderer, M. Ohl, D. Richter, Phys. Rev. Lett. 110, 108303 (2013)

METHODS



C 1 Scattering

R. Zorn
Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Scattering from static systems **3**
2.1 Structure factor from density 3
2.2 Structure factor from pair correlation function 5
2.3 Example: form factor of a sphere 7
2.4 Example: liquid structure factor 9
2.5 Structure factor and form factor 10

3 Scattering from dynamic systems **11**
3.1 Example: diffusion 17

4 Scattering probes: photons, neutrons, electrons **19**

5 Scattering instruments for soft matter investigations **23**
5.1 Light scattering 23
5.2 X-ray scattering 25
5.3 Neutron scattering 27

A Elementary properties of the Fourier transform **31**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Scattering methods play an important rôle in the investigation of soft matter systems. In contrast to real space (microscopy) methods, the result of a scattering experiment is related to the Fourier transform of the structure and dynamics of the system. This is why scattering methods are often said to work in 'reciprocal space'. This has certain advantages and disadvantages. While the main disadvantage – the absence of an immediately comprehensible three-dimensional 'picture' of the system – is immediately clear, the advantages need some explanation:

First of all, the direct-space observation of soft matter objects on a molecular level is often impossible due to technical reasons. Their size is often too small for light microscopy. Electron microscopy may be impossible in their native environment (solution, melt). Atomic force microscopy may be difficult because of their softness. In these situations scattering methods are the only way to obtain structural information[1]. But even if microscopy methods are available scattering methods have the advantage of intrinsic averaging. Average properties (e.g. the average molecular weight of polymer molecules) can be obtained without doing a statistics of individual observations.

In the literature, a general overview of scattering methods focused on soft matter applications is rarely given, maybe with the exception of ref. 1 and (for polymers) ref. 2. Most textbooks focus on individual methods without specifying the application further: neutron scattering [3–7], light scattering [8–12] and x-ray scattering [13–15].

The central statement of this lecture is that scattering experiments indirectly measure correlation functions. The usual derivation of scattering laws is based on the fact that the scattering law (for neutrons, photons or any other radiation) is essentially the absolute square of the Fourier transform of a scattering density. In Fig. 1 this is shown as the left way from the density $\rho(\mathbf{r})$ to $S(Q)$. The Wiener-Khintchine theorem

$$|\mathcal{F}[f(x)]|^2 = \mathcal{F}[\langle f(0)f(x)\rangle] \tag{1}$$

(with the Fourier transform $\mathcal{F}$ defined as in the appendix) now states that the absolute square of a Fourier transform is the Fourier transform of the autocorrelation function. This opens another way (the right one in Fig. 1) to calculate the scattering law. Apart from elucidating the meaning of the scattering law in another way, this gives an alternative to calculate it even if the density itself is not known.

This lecture will in the first section treat the results from a static system where the scattering is completely elastic. In this situation the scattering will only contain information about the structure. Strictly speaking, this is a fictitious assumption because all materials show some dynamics (quantum-mechanically even at zero temperature). Nevertheless, the broad range of diffraction methods is covered with sufficient accuracy. The next part of the lecture will deal with inelastic scattering. In this experiment, scattering gives information about the structure via the momentum transfer and about the dynamics via the energy transfer. For inelastic scattering a Fourier transform in time has to be carried out in addition leading from the time correlation function to the scattering function. Finally, the differences between probe beams (light, x-rays, neutrons) will be explained and a selection of scattering instruments for soft matter investigations will be presented.

[1] In this lecture "structure" will be understood as the exact positions of atoms in space. Of course, in a more loose sense also methods as NMR and even spectroscopic methods give information on the structure.

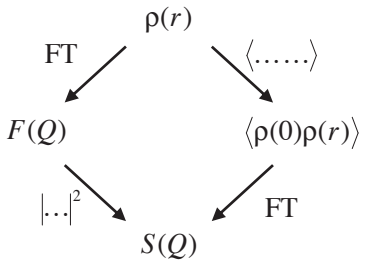


Fig. 1: *The two ways to calculate the scattering law from the microscopic density, left: as the absolute-squared Fourier transform of the density, right: as the Fourier transform of the correlation function.*

2 Scattering from static systems

In this section it will be assumed that the scattering system is *static*. It is either represented by fixed positions of point scatterers in space, $\mathbf{r}_j$, or a time-independent density, $\rho(\mathbf{r})$. The former case can be included in the latter by considering the microscopic density as a sum of delta functions:

$$\rho(\mathbf{r}) = \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{r}_j)\,. \tag{2}$$

From the fact that the scatterers are fixed follows that the scattering will be *elastic*, i.e. the energy of the scattered particles will not change due to the scattering process. This is clear from classical mechanics because a system which is static before and after the scattering process cannot exchange energy. The equivalent argument from the wave picture would be that upon scattering by fixed centres there is no Doppler shift of the frequency.

2.1 Structure factor from density

The result of an elastic scattering experiment is usually expressed in terms of the *differential cross-section* which is the probability density that a particle is scattered into a solid angle element $\mathrm{d}\Omega$ normalised to the intensity of the incident beam:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left\langle \left| \sum_{j=1}^{N} b_j \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_j) \right|^2 \right\rangle . \tag{3}$$

b_j is a measure of the 'scattering power' of the particle. From the dimensions of the other quantities it is obvious that it has the dimension [length]. Therefore, b_j is called the *scattering length*. Note that the scattering length is not necessarily positive. $b_j < 0$ just means that scattering leads to a reversal of the amplitude, in other words a phase shift π. The scattering length may even be complex. In that case, the imaginary part corresponds to absorption of the scattered particle by the scatterer.

It can be seen that expression (3) does not contain the scattering angle 2θ directly but a *scattering vector* $\mathbf{Q}$. It is the vectorial difference of the wave vector $\mathbf{k}'$ after scattering and that before scattering, $\mathbf{k}$. The wave vectors are defined by having the length $|\mathbf{k}| = k = 2\pi/\lambda$ and

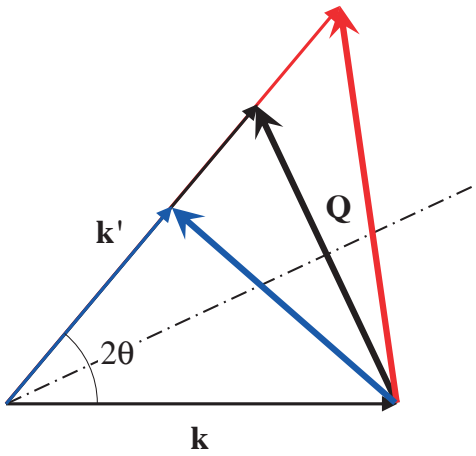


Fig. 2: *Definition of the scattering vector* **Q** *in terms of the incident and final wave vectors* **k** *and* **k**′*. The black (isosceles) triangle corresponds to elastic scattering. The blue and red ones correspond to inelastic scattering with energy loss or gain of the scattered radiation, respectively.*

the direction of the propagation of the wave. For elastic scattering $k' = k$, and the definition of **Q** is graphically demonstrated by the black (isosceles) triangle in Fig. 2 resulting in

$$Q = \frac{4\pi}{\lambda} \sin\theta \,. \tag{4}$$

From this equation one can see that scattering depends on a combination of the scattering angle and the wavelength of the scattered radiation. The same Q can be obtained by different combinations of 2θ and λ.
At this point it is necessary to explain the meaning of the average $\langle \ldots \rangle$ in (3) and justify it. Of course for a completely arrested system and completely coherent radiation, (3) would be valid without the average. Experimentally, this situation is only realised in laser light scattering from rigid objects. There, the experiments as well as the calculation do not yield a smooth function $\mathrm{d}\sigma/\mathrm{d}\Omega$ but an assembly of so-called speckles. For two reasons this situation is exceptional and the observed scattering is actually an average:

1. If a dynamics exists, even if it is sufficiently slow not to cause a noticeable inelasticity, the particles will rearrange over the duration of the experiment. In this sense, $\langle \ldots \rangle$ expresses a temporal average over the experimental time.

2. If the radiation used is not highly coherent, the sum over the amplitudes in (3) has to be restricted to the coherence volume which is usually much smaller than the sample volume. The results from the individual regions have to be added as intensities, i.e. after the absolute-square. This implies the same average but to be interpreted as a thermodynamic average over different realisations of the particle positions. In the case of ergodic systems both averages have the same result.

With the assumption that all scatterers are identical (for neutron scattering implying that they are the same isotope and have the same spin orientation) one can factor out the material-specific

properties N and b:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = |b|^2 N S(\mathbf{Q})\,. \tag{5}$$

with the remaining term

$$S(\mathbf{Q}) = \frac{1}{N}\left\langle \left| \sum_{j=1}^{N} \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_j) \right|^2 \right\rangle \tag{6}$$

which depends solely on the statistics of the positions of the scatterers. $S(\mathbf{Q})$ is called *structure factor*.

If the scattering is not effected by individual scatterers but by a field (e.g. the magnetic field for neutrons) or a distribution (e.g. the electron density for x-rays) one has to use a continuum description instead of (6):

$$S(\mathbf{Q}) = \frac{1}{N}\left\langle \left| \int_V \mathrm{d}^3 r \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r})\rho(\mathbf{r}) \right|^2 \right\rangle\,. \tag{7}$$

It is easy to verify that this expression corresponds to (6) with the definition (2) of the microscopic density inserted. Expression (7) is the absolute square of the Fourier transform of the density and thus represents the 'left way' in Fig. 1.

But even if the individual scatterers are point-like, the continuum description may be useful if their exact positions are not known but only their mesoscopic densities. This is often the case for soft matter systems. Then, expression (7) will be a good approximation as long as the length scale defined by Q is large compared to the distances between the scatterers, $Q \ll 2\pi/\text{distance}$ (e.g. for small-angle x-ray or -neutron scattering).

At that point a simple way to introduce mixed scatterers is to start with the *scattering length density*

$$\rho_b(\mathbf{r}) = \sum_{j=1}^{N} b_j \delta(\mathbf{r}-\mathbf{r}_j)\,. \tag{8}$$

instead of the density. By including the scattering properties in the density, equation (5) can be written as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left\langle \left| \int_V \mathrm{d}^3 r \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r})\rho_b(\mathbf{r}) \right|^2 \right\rangle\,. \tag{9}$$

Thus, the differential cross section is the absolute square of the Fourier transform of the scattering length density. For neutron scattering, this concept is used to obtain a low-resolution description for small-angle scattering and reflectometry. For light scattering the local dielectric constant of the medium plays the rôle of $\rho_b(\mathbf{r})$.

2.2 Structure factor from pair correlation function

The second way to derive the scattering law starts with applying the definition of the absolute square, $|X|^2 = X^* X$ to equation (6):

$$\begin{aligned} S(\mathbf{Q}) &= \frac{1}{N}\left\langle \left(\sum_{j=1}^{N} \exp(-\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_j) \right) \left(\sum_{k=1}^{N} \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_k) \right) \right\rangle \\ &= \frac{1}{N}\sum_{j,k=1}^{N} \left\langle \exp(\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k-\mathbf{r}_j)) \right\rangle\,. \end{aligned} \tag{10}$$

From this expression two characteristic properties of scattering become clear:

1. The scattering law arises from particle pairs (j, k).
2. Only distances between particles enter the expression, not the individual positions. The scattering law remains invariant under translation of the whole sample.

Analogously to $S(\mathbf{Q})$ the differential scattering cross section can be expressed in terms of particle distances as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left\langle \sum_{j,k=1}^{N} {b_j}^* b_k \exp\left(\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k - \mathbf{r}_j)\right) \right\rangle . \tag{11}$$

In order to proceed in a similar way as before, we introduce the two-particle density

$$\rho(\mathbf{r}_1)\rho(\mathbf{r}_2) = \sum_{j,k=1}^{N} \delta(\mathbf{r}_1 - \mathbf{r}_j)\delta(\mathbf{r}_2 - \mathbf{r}_k) \tag{12}$$

which is the joint probability that particle j is found at $\mathbf{r}_1$ and particle k at $\mathbf{r}_2$. It is important that in general the average of this probability density is not just the product of the average densities:

$$\langle\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle \neq \langle\rho(\mathbf{r}_1)\rangle\langle\rho(\mathbf{r}_2)\rangle = {\rho_0}^2 \tag{13}$$

($\rho_0 = N/V$). The reason for this is that usually there is an interaction between particles which enhances or reduces the probability for particles close to each other. E.g. if one imagines particles with a hard core of radius R then $\langle\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle$ vanishes for all $\mathbf{r}_1$ and $\mathbf{r}_2$ which would imply a 'collision' of the particles, $0 < |\mathbf{r}_2 - \mathbf{r}_1| < 2R$. But still, in a translationally invariant system one of the positions can be chosen arbitrarily, especially as the origin, so that

$$\langle\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle = \langle\rho(\mathbf{0})\rho(\mathbf{r}_2 - \mathbf{r}_1)\rangle = \rho_0 \left\langle \sum_{j,k=1}^{N} \delta(\mathbf{r}_j - \mathbf{r}_k + \mathbf{r}_2 - \mathbf{r}_1) \right\rangle . \tag{14}$$

For a system of identical scatterers the two-particle density (12) can now be used to express the structure factor:

$$\begin{aligned} S(\mathbf{Q}) &= \frac{1}{N}\left\langle \int_V \mathrm{d}^3 r_1 \int_V \mathrm{d}^3 r_2 \exp(\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_2 - \mathbf{r}_1))\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\right\rangle \\ &= \frac{1}{\rho_0}\int_{V_d} \mathrm{d}^3 r \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r})\langle\rho(\mathbf{0})\rho(\mathbf{r})\rangle . \end{aligned} \tag{15}$$

Note that in this last expression $\mathbf{r}$ does not have the meaning of an absolute position but that of a vectorial distance and consequently the volume of integration V_d is not the sample volume V but the volume of possible distances within the sample.
In the literature often alternative definitions of the pair correlation function are used (instead of using $\langle\rho(\mathbf{0})\rho(\mathbf{r})\rangle$ directly). The most common definition in the context of soft matter is

$$g(\mathbf{r}) = \frac{\langle\rho(\mathbf{0})\rho(\mathbf{r})\rangle}{{\rho_0}^2} - \frac{\delta(\mathbf{r})}{\rho_0} . \tag{16}$$

The normalisation by ${\rho_0}^2$ has the effect that for non-interacting particles or at distances where the interaction is weak, $g(\mathbf{r}) = 1$. The subtraction of the delta function removes the singularity

of $\langle\rho(\mathbf{0})\rho(\mathbf{r})\rangle$ at $\mathbf{r} = \mathbf{0}$ due to the $j = k$ terms in (12). With this pair correlation function the structure factor can be written as

$$S(\mathbf{Q}) = 1 + \rho_0 \int_{V_d} \mathrm{d}^3 r \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r})(g(\mathbf{r}) - 1)\,. \tag{17}$$

Here, the $1+$ compensates the delta function term subtracted in (16). In addition, one usually writes $g(\mathbf{r}) - 1$ instead of simply $g(\mathbf{r})$ in the Fourier transform. This avoids a delta function term arising in the limit $V_d \to \infty$ at $Q = 0$. In that limit, this 'trick' only changes the result at $Q = 0$ which is the (unobservable) forward scattering. Nevertheless, strictly speaking, one loses the scattering contribution by the overall sample shape. But this only affects the very low Q region if the sample has macroscopic dimensions $\gg 2\pi/Q$.
In many physical systems the interaction between particles is not directional with the consequence that $g(\mathbf{r})$ depends only on the distance $r = |\mathbf{r}|$. In this case by symmetry follows that also $S(\mathbf{Q})$ is only a function of $Q = |\mathbf{Q}|$. Formula (103) from the appendix can be applied resulting in

$$S(Q) = 1 + \frac{4\pi\rho_0}{Q} \int_0^\infty (g(r) - 1)\sin(Qr) r \mathrm{d}r\,. \tag{18}$$

2.3 Example: form factor of a sphere

On a length scale larger than the distance of the individual scatterers, the density of a sphere of radius R is given as

$$\rho(r) = \begin{cases} \rho_0 & \text{for } r < R \\ 0 & \text{for } r > R \end{cases}\,. \tag{19}$$

Because of the spherical symmetry one can convert the volume integral in (7) into a one-dimensional integral using (103) from the appendix:

$$\begin{aligned} \int_V \mathrm{d}^3 r \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r})\rho(\mathbf{r}) &= \frac{4\pi}{Q}\int_0^\infty \rho(r) r \sin(Qr)\mathrm{d}r \\ &= \frac{4\pi\rho_0}{Q^3}\left(\sin(QR) - QR\cos(QR)\right)\,. \end{aligned}$$

With the relation between the density of the scatterers and their number and the volume of the sphere, the prefactor becomes $3N/Q^3R^3$, and inserting the result in (7) yields

$$S(Q) = \frac{9N}{Q^3R^3}\left(\sin(QR) - QR\cos(QR)\right)^2\,. \tag{20}$$

One can see that the *structure* factor of the sphere would be proportional to the number of scatterers it contains. In practice, for a single particle, one prefers a normalisation which does not depend on the number of scatterers within a particle. This leads to the definition of the *form* factor, where $1/N$ in equation (7) is replaced by $1/N^2$. For the sphere the form factor is

$$P(Q) = \frac{9}{Q^6R^6}\left(\sin(QR) - QR\cos(QR)\right)^2 \tag{21}$$

shown in Fig. 3. The zeros predicted by (21) are usually not found in the experiment but only more-or-less pronounced minima. The reason is that experimental samples usually consist

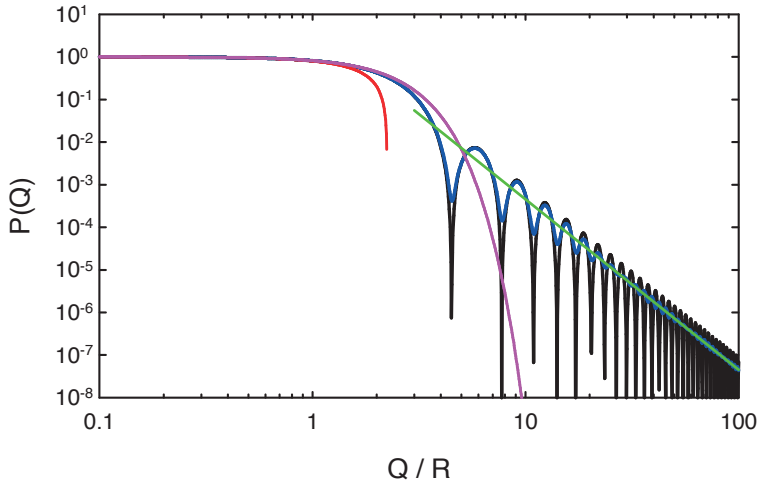


Fig. 3: *Form factor of a sphere (black), average for spheres distributed equally over* $R = 0.95 \ldots 1.05$ *(blue), low* Q *expansion (22) (red), Guinier approximation (23) (magenta), Porod limit (25) (green).*

of spheres of slightly different radii (polydispersity). This leads to a smearing out which is simulated by a $\pm 5\%$ variation in R in the blue curve of Fig. 3.
A series expansion of (21) yields

$$P(Q) = 1 - \frac{Q^2 R^2}{5} + \mathcal{O}(Q^4)\,. \tag{22}$$

The Q^2 term in the expansion (22) also has a more general significance. It is actually related to the fact that for *any* form factor to second order in Q the *Guinier approximation* holds:

$$P(Q) \approx \exp\left(-\frac{Q^2 {R_\mathrm{g}}^2}{3}\right) \tag{23}$$

where

$${R_\mathrm{g}}^2 = \frac{\int_V r^2 \rho(r) \mathrm{d}^3 r}{\int_V \rho(r) \mathrm{d}^3 r} = \frac{1}{2} \frac{\int_{V_d} r^2 g(r) \mathrm{d}^3 r}{\int_{V_d} g(r) \mathrm{d}^3 r} \tag{24}$$

is the *radius of gyration* of the particle[2]. By comparison of (22) and (23) or direct calculation from (24) results $R_\mathrm{g} = \sqrt{3/5} R$. As Fig. 3 shows, the Guinier approximation approximates the sphere from factor slightly better than the second order approximation (22) because it contains higher order terms in Q despite with incorrect coefficients.
Finally, a general high Q property of the scattering can be demonstrated with the example of a sphere. For large Q the $\sin(QR)$ term in (21) becomes negligible compared to $QR\cos(QR)$. Also, assuming some polydispersity or finite instrumental resolution which are always present in a real experiment, $(\cos(QR))^2$ can be replaced by its average $\overline{\cos^2 x} = 1/2$ and $\cos(QR)$ $\sin(QR)$ by $\overline{\cos x \sin x} = 0$, so that:

$$\overline{\lim_{Q\to\infty}} P(Q) = \frac{9}{2} \frac{1}{Q^4 R^4}\,. \tag{25}$$

[2]In the first expression it is essential that r is measured with respect to the center of gravity of the particle.

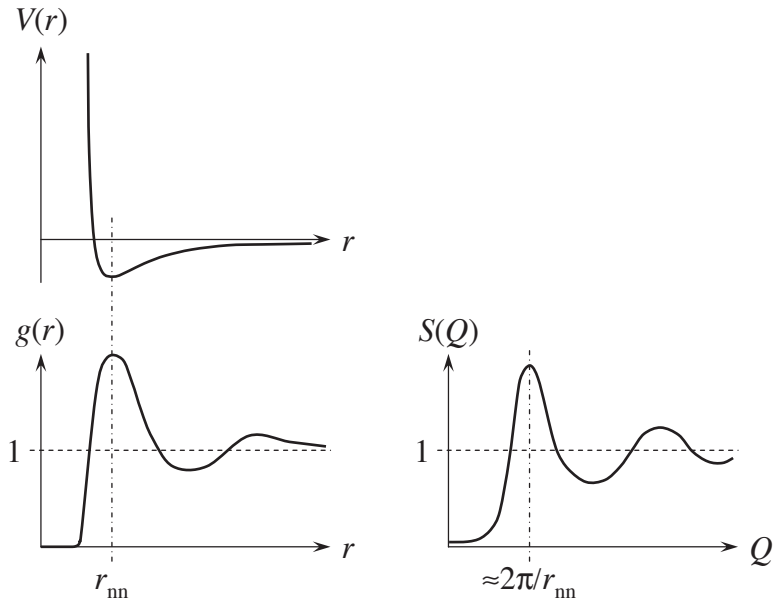


Fig. 4: *Schematic representation of interaction potential $V(r)$, pair correlation function $g(r)$, and scattering function $S(Q)$.*

Again, this is the special case of a general relation, the *Porod law*, which relates the high Q scattering to the ratio surface area S_p / volume V_p:

$$\overline{\lim_{Q\to\infty}} P(Q) = \frac{2\pi S_\mathrm{p}}{{V_\mathrm{p}}^2}\frac{1}{Q^4}\,. \tag{26}$$

2.4 Example: liquid structure factor

This second example is not an exact calculation but more a rough description of the features to be expected for scattering from a liquid (Fig. 4). A liquid also does not fulfil the requirement that the structure is static in the strict sense required above. Nevertheless, as will be derived in the section 3, a diffraction experiment will yield an $S(Q)$ corresponding to the instantaneous structure. But what is more a problem for the mathematical treatment is that for a given interparticle potential $V(r)$ there is no exact way to derive the pair correlation function $g(r)$. There are only approximative analytical methods [16] and numerical methods available for this purpose. Nevertheless it can be expected that there is a preferential nearest-neighbour distance r_nn which is roughly defined by the minimum of the interparticle potential and corresponds to a maximum in $g(r)$. As explained before $g(r)$ will drop sharply for too short distances because of the strong repulsion. For large r there will be no significant interaction between the particles so that the joint probability $\langle\rho(\mathbf{0})\rho(\mathbf{r})\rangle$ will become the product of the average densities ${\rho_0}^2$ and in consequence $\lim_{r\to\infty} g(r) = 1$.

From $g(r)$ by use of equation (18) the structure factor can be calculated. Although again an exact result cannot be given, several general features can be stated: For $Q \to \infty$, $\exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r})$

becomes a rapidly oscillating function and the integral vanishes. Then one has

$$\lim_{Q\to\infty} S(Q) = 1\,. \tag{27}$$

For $Q \to 0$, $S(Q)$ measures only the overall density fluctuation, i.e. the fluctuation of the particle number:

$$\lim_{Q\to 0} S(Q) = \frac{V^2\langle\delta\rho^2\rangle}{N} = \frac{\langle N^2\rangle - \langle N\rangle^2}{\langle N\rangle} = \rho_0 k_\mathrm{B} T \kappa_T\,. \tag{28}$$

Here, k_B denotes the Boltzmann constant, T the temperature and κ_T the isothermal compressibility. At intermediate Q, the structure factor of liquids shows a diminishing series of broad peaks, remainders of the Bragg peaks of a crystalline structure. The first peak occurs at a scattering vector roughly corresponding to the next neighbour distance by $Q_\mathrm{max} = 2\pi/r_\mathrm{nn}$.

2.5 Structure factor and form factor

Many physical systems are not constituted by point scatterers but nevertheless the density is not just some irregular function of coordinate. Rather, the scatterers are structured particles and this structure is identically repeated at different centres in space. Mathematically, this can be captured as a convolution:

$$\rho(\mathbf{r}) = \rho_\mathrm{centres}(\mathbf{r}) \otimes \rho_\mathrm{particle}(\mathbf{r}) = \int \mathrm{d}^3 r' \rho_\mathrm{centres}(\mathbf{r}')\rho_\mathrm{particle}(\mathbf{r}-\mathbf{r}') \tag{29}$$

where $\rho_\mathrm{centres}(\mathbf{r}) = \sum_j \delta(\mathbf{r} - \mathbf{r}_j^\mathrm{centres})$.
Now the convolution theorem (115) states that the Fourier transform of the density (29) is the product of the Fourier transforms of the underlying densities:

$$\mathcal{F}[\rho(\mathbf{r})] = \mathcal{F}[\rho_\mathrm{centres}(\mathbf{r})]\mathcal{F}[\rho_\mathrm{particle}(\mathbf{r})]\,. \tag{30}$$

Consequently, the scattering of the whole system, which for distinction should be denoted $I(Q)$ here, is

$$I(Q) = S(Q)P(Q) \tag{31}$$

where $S(Q)$ is the structure factor of the centre positions alone and $P(Q)$ is the form factor of the density within a single particle[3].
The form factor is a rapidly decaying function, with a range of roughly $Q_\mathrm{max} \approx 2\pi/d$ where d is the size of the scatterer. In case of x-ray scattering from atoms d is in the Ångström range; therefore, $P(Q)$ decays within a couple of $\mathrm{Å}^{-1}$. This implies that $S(Q)$ is damped significantly by (31) and higher order Bragg peaks in crystals or short range correlations in liquids cannot be observed well. For neutron scattering (nuclear, not magnetic) a potential with just a few femtometres range is relevant. There, the form factor would only decay at extremely high Q and is practically constant for all relevant Q values. Nevertheless, also for neutron scattering the form factor/strcture factor dichotomy is useful in the case of complex liquids (colloids,

[3]Note that in order to get equation (31) literally correct, one needs the different normalisation of $P(Q)$ mentioned in example 2.3, namely by $1/N^2$ instead of $1/N$. The exact way of normalising the terms in the product (31) is non-uniform in the literature, so that the formula may appear with different additional prefactors, e.g. the average density within the particle.

microemulsions, polymer solutions etc.). In that case, one has an arragement of statistically equal objects of nanometre size leading to a form factor decaying at a Q of the order of 0.1 Å^{-1}. There is often some confusion in the use of "form factor" and "structure factor". The reason for this is that the two terms are defined relatively to each other. Imagine for example x-ray scattering from a polymer solution. The solution is constituted by polymer molecules which itself are constituted by atoms. Thus, there is a three-level hierarchy. With respect to the levels molecule-atom we have an atomic form factor $P_{\mathrm{at}}(Q)$ and a polymer molecule structure factor $S_{\mathrm{mol}}(Q)$. Looking at the levels solution-molecule, there is a molecular form factor $P_{\mathrm{mol}}(Q)$ and a solution structure factor $S_{\mathrm{sol}}(Q)$. But in the end $S_{\mathrm{mol}}(Q) = P_{\mathrm{mol}}(Q)$, meaning that the question whether the polymer molecule has a structure factor or a form factor depends on the view of this intermediate level.

3 Scattering from dynamic systems

Here, the more realistic situation will be considered in which the particles of the sample are moving. Their dynamics will be described by *trajectories* $\mathbf{r}_j(t)$ which implies that the result is only valid in classical approximation. This assumption is usually uncritical for soft matter systems because experiments are done at high temperatures where quantum mechanical effects are unobservable. In most scattering experiments the motion of the particles is observed indirectly via the inelasticity, i.e. the energy transfer $\hbar\omega = E' - E$ of the scattering process. The notable exception is dynamic light scattering where a time correlation function is measured directly (subsection 5.1).

In analogy to (3) the *double differential cross-section* is defined as the probability density that a neutron is scattered into a solid angle element $\mathrm{d}\Omega$ with an energy transfer $\hbar\omega \ldots \hbar(\omega + \mathrm{d}\omega)$. Within a certain approximation, which is usually valid for soft matter systems, it can be given as[4]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega \mathrm{d}\omega} = \frac{1}{2\pi}\frac{k'}{k}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t \left\langle \sum_{j,k=1}^{N} b_j^* b_k \exp\left(\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k(t) - \mathbf{r}_j(0))\right)\right\rangle . \tag{32}$$

It can be seen that the 'indirectness' of scattering leads to another Fourier transform in time $t \to \omega$ (in addition to that in space $r \to Q$). A factor (which was not visible before because $k' = k$ for scattering from static systems) is the ratio of the wave numbers k'/k. This factor can be easily understood for massive particles as neutrons as the ratio of velocities v'/v which modulates the intensity of the scattered beam.

Another place where $k' \neq k$ has to be taken into account is that Q now does not anymore result from the isosceles construction in Fig. 2 drafted in black but from scattering triangles as those in blue and red. Application of the cosine theorem leads to the following expression for Q in the inelastic situation:

$$\begin{aligned} Q &= \sqrt{k^2 + k'^2 - 2kk'\cos(2\theta)} \\ &= \sqrt{\frac{8\pi^2}{\lambda^2} + \frac{2m\omega}{\hbar} - \frac{4\pi}{\lambda}\sqrt{\frac{4\pi^2}{\lambda^2} + \frac{2m\omega}{\hbar}}\cos(2\theta)} \quad \text{for neutrons.} \end{aligned} \tag{33}$$

[4] A correct derivation of the double differential cross section for inelastic scattering is only possible by quantum mechanical means. This is usually done in textbooks on neutron scattering as refs. 4, 5. It would be beyond the scope of this lecture to present this derivation in detail.

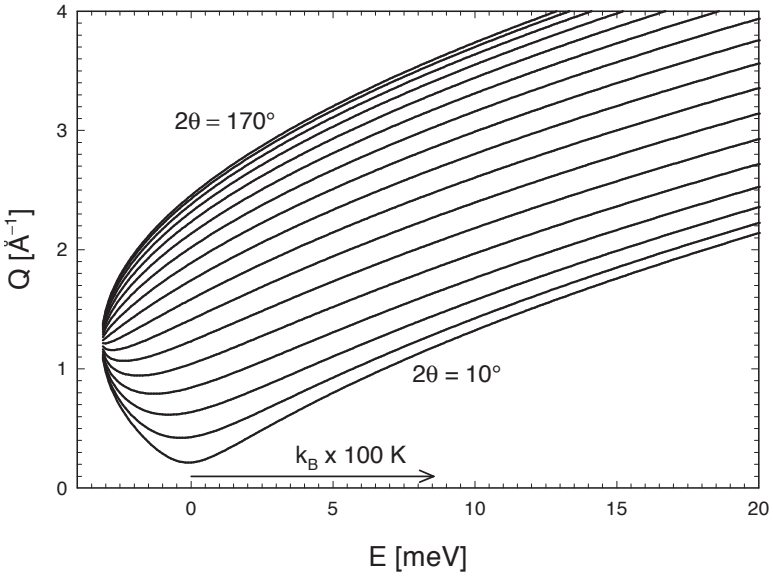


Fig. 5: *Scattering vectors Q accessed by a neutron scattering experiment with the detector at scattering angles $2\theta = 10 \ldots 170°$ vs. the energy transfer $\hbar\omega$ (incident wavelength $\lambda = 0.51$ nm). For comparison the thermal energy $k_\mathrm{B}T$ corresponding to 100 K is indicated by an arrow.*

Especially, it has to be observed now that Q also depends on $\hbar\omega$ implying that Q is not anymore constant for a single scattering angle. Fig. 5 shows the magnitude of this effect for typical parameters of a neutron scattering experiment. It can be seen that it is by no means negligible for typical thermal energies of the sample even at temperatures as low as 100 K.

Expression (32) is valid for (nuclear, non-polarised) neutron scattering and the most simple if compared to light scattering or x-ray scattering. The expressions for the other scattering probes would include further terms involving the polarisation of the beam and the polarisability of the sample. Because these are not immediately relevant for soft matter systems, we will continue with this simplest formulation of inelastic scattering.

In order to derive a quantity similar to the structure factor (17), one assumes again a system of N *chemically* identical particles. But in order to capture the feature of incoherent scattering, present in neutron scattering, it is assumed that these particles do not have identical scattering lengths but individual randomly distributed scattering lengths with the average $\overline{b} = (1/N)\sum_j b_j$ and the variance $\overline{|b|^2} - \left|\overline{b}\right|^2 = \overline{\left|b - \overline{b}\right|^2} = (1/N)\sum_j \left|b_j - \overline{b}\right|^2$. The most obvious reason for the variance of scattering lengths is that chemically identical atoms may be different isotopes. Because the neutron scattering length is a nuclear property it may differ from isotope to isotope. But even in monisotopic systems there may be such a variance due to disorder of the nuclear spin orientations because the scattering length also depends on the combined spin state of the scattered neutron and the scattering nucleus. It is more difficult to see how incoherent scattering may arise in light scattering. There the individual scattering centres may be colloidal particles with an effective b depending on the size and the refractive index. The analogue to isotope disorder would be here to have particles with identical size and interaction but different refractive indices. This trick is exploited to obtain incoherent scattering from colloidal systems [17] and

thereby gain access to the self-correlation.
The sum in expression (32) can be decomposed into one over different indices and one over identical indices,

$$\sum_{j,k=1}^{N} b_j^* b_k \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k(t)-\mathbf{r}_j(0))} = \sum_{j\neq k=1}^{N} b_j^* b_k \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k(t)-\mathbf{r}_j(0))} + \sum_{j=1}^{N} |b_j|^2 \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_j(t)-\mathbf{r}_j(0))}\,, \tag{34}$$

which have to be averaged in a different way with respect to the distribution of scattering lengths. In the first term b_j^* and b_k can be averaged separately because the different particle scattering lengths are uncorrelated: $\overline{b^*}\,\overline{b} = \overline{b}^*\overline{b} = |\overline{b}|^2$. In the second term one has to average *after* taking the absolute square:

$$= \sum_{j\neq k=1}^{N} |\overline{b}|^2 \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k(t)-\mathbf{r}_j(0))} + \sum_{j=1}^{N} \overline{|b|^2} \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_j(t)-\mathbf{r}_j(0))}\,. \tag{35}$$

In order to avoid the sum over distinct particles, the first sum is complemented by the $j = k$ terms, $|\overline{b}|^2 \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_j(t)-\mathbf{r}_j(0))}$, and to compensate, these terms are subtracted in the second sum:

$$= \sum_{j,k=1}^{N} |\overline{b}|^2 \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k(t)-\mathbf{r}_j(0))} + \sum_{j=1}^{N} \left(\overline{|b|^2} - |\overline{b}|^2\right) \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_j(t)-\mathbf{r}_j(0))}\,. \tag{36}$$

With this result it is possible to express the double differential cross section as

$$\frac{\partial\sigma}{\partial\Omega\partial\omega} = N\frac{k'}{k}\left(\left|\overline{b}\right|^2 S_{\mathrm{coh}}(\mathbf{Q},\omega) + \left(\overline{|b|^2} - \left|\overline{b}\right|^2\right) S_{\mathrm{inc}}(\mathbf{Q},\omega)\right) \tag{37}$$

with

$$S_{\mathrm{coh}}(\mathbf{Q},\omega) = \frac{1}{2\pi N}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t \sum_{j,k=1}^{N} \left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k(t)-\mathbf{r}_j(0))}\right\rangle \tag{38}$$

and

$$S_{\mathrm{inc}}(\mathbf{Q},\omega) = \frac{1}{2\pi N}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t \sum_{j=1}^{N} \left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_j(t)-\mathbf{r}_j(0))}\right\rangle\,. \tag{39}$$

The quantities defined by (38) and (39) are called *coherent* and *incoherent scattering function* or *dynamic structure factors*.
The prefactors of the scattering functions in expression (37) are often expressed by the scattering cross sections

$$\sigma_{\mathrm{coh}} = 4\pi\left|\overline{b}\right|^2, \quad \sigma_{\mathrm{inc}} = 4\pi\left(\overline{|b|^2} - \left|\overline{b}\right|^2\right) \tag{40}$$

which (for the incoherent part in general and for the coherent in the limit $Q \to \infty$) give the scattering into all directions, i.e. the solid angle 4π.
In some cases it is interesting to consider the part of expression (38) before the time-frequency Fourier transform, called *intermediate coherent scattering function*:

$$I_{\mathrm{coh}}(\mathbf{Q},t) = \frac{1}{N}\sum_{jk}\left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k(t)-\mathbf{r}_j(0))}\right\rangle\,. \tag{41}$$

Its value for $t = 0$ expresses the correlation between atoms *at equal times*. A theorem on Fourier transforms (equation (112) of the appendix) tells that this is identical to the integral of the scattering function over all energy transfers:

$$I_{\mathrm{coh}}(\mathbf{Q},0) = \frac{1}{N}\sum_{jk}\left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k-\mathbf{r}_j)}\right\rangle = S(\mathbf{Q}) = \int_{-\infty}^{\infty} S_{\mathrm{coh}}(\mathbf{Q},\omega)\mathrm{d}\omega\,. \tag{42}$$

The concrete significance of this relation is that a diffraction experiment, which does not discriminate energies and thus implicitly integrates over all $\hbar\omega$, only shows the instantaneous correlation of the atoms, viz the structure of the sample[5]. $S(\mathbf{Q})$ is the structure factor as derived in section 2 for the static situation. The dynamic information is lost in the integration process.
Similarly the *incoherent intermediate scattering function* is

$$I_{\mathrm{inc}}(\mathbf{Q},t) = \frac{1}{N}\sum_{j=1}^{N}\left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_j(t)-\mathbf{r}_j(0))}\right\rangle \tag{43}$$

with

$$I_{\mathrm{inc}}(\mathbf{Q},0) = \frac{1}{N}\sum_{j=1}^{N}\left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_j-\mathbf{r}_j)}\right\rangle = 1 = \int_{-\infty}^{\infty} S_{\mathrm{inc}}(\mathbf{Q},\omega)\mathrm{d}\omega\,. \tag{44}$$

Note that this result is independent of the actual structure of the sample. Also it is the same result as the high Q limit $S(Q) \to 1$ (27). This is a consequence of the more general fact that coherent and incoherent scattering become indistinguishable for large Q. Integration of the double-differential cross section (37) over ω shows that also the static scattering contains an incoherent contribution. But because of (44), this term is constant in Q. It constitutes a flat background in addition to the $S(Q)$-dependent scattering. In some cases (e.g. small-angle scattering) it may be necessary to correct for this, in other cases (e.g. diffraction with polarisation analysis) it may even be helpful to normalise the coherent scattering.
In the paragraphs before it was shown that the value of the intermediate scattering functions at $t = 0$ corresponds to the integral of the scattering function over an infinite interval. This is a consequence of a general property of the Fourier transform. There is also the inverse relation that the value of $S(\mathbf{Q},\omega)$ at $\omega = 0$ is related to the integral of $I(\mathbf{Q},t)$ over all times (equation (113)). The most important case is here when $I(\mathbf{Q},t)$ does not decay to zero for infinite time but to a finite value $f(\mathbf{Q})$. In that case the integral is infinite implying that $S(\mathbf{Q},\omega)$ has a delta function contribution at $\omega = 0$). This means that the scattering contains a strictly elastic component. Its strength can be calculated by decomposing the intermediate scattering function into a completely decaying part and a constant for the coherent and the incoherent scattering:

$$I_{[\mathrm{coh|inc}]}(\mathbf{Q},t) = I^{\mathrm{inel}}_{[\mathrm{coh|inc}]}(\mathbf{Q},t) + f_{[\mathrm{coh|inc}]}(\mathbf{Q})\,. \tag{45}$$

Because the Fourier transform of constant one is the delta function this corresponds to

$$S_{[\mathrm{coh|inc}]}(\mathbf{Q},\omega) = S^{\mathrm{inel}}_{[\mathrm{coh|inc}]}(\mathbf{Q},\omega) + S^{\mathrm{el}}_{[\mathrm{coh|inc}]}(\mathbf{Q})\delta(\omega)\,. \tag{46}$$

[5] Strictly speaking, this is only an approximation. There are several reasons why the integration in the diffraction experiment is not the 'mathematical' one of (42): (1) On the instrument the integral is taken along a curve of constant 2θ in Fig. 5 while constant Q would correspond to a horizontal line. (2) The double differential cross-section (37) contains a factor k'/k which depends on ω via (33). (3) The detector may have an efficiency depending on wavelength which will introduce another ω-dependent weight in the experimental integration. All these effects can be taken into account in the so-called Placzek corrections [18–20].

where $S^{\mathrm{el}}_{[\mathrm{coh|inc}]}(\mathbf{Q}) = f_{[\mathrm{coh|inc}]}(\mathbf{Q})$, the *elastic coherent/incoherent structure factor*, can be written as

$$S^{\mathrm{el}}_{\mathrm{coh}}(\mathbf{Q}) = \frac{1}{N}\sum_{j,k=1}^{N}\left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_k(\infty)-\mathbf{r}_j(0))}\right\rangle , \tag{47}$$

$$S^{\mathrm{el}}_{\mathrm{inc}}(\mathbf{Q}) = \frac{1}{N}\sum_{j=1}^{N}\left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_j(\infty)-\mathbf{r}_j(0))}\right\rangle . \tag{48}$$

Here, $t = \infty$ indicates a time which is sufficiently long that the correlation with the position at $t = 0$ is lost. For the elastic incoherent structure factor (EISF) this lack of correlation implies that the terms with initial and final positions can be averaged separately:

$$\begin{aligned} S^{\mathrm{el}}_{\mathrm{inc}}(\mathbf{Q}) &= \frac{1}{N}\sum_{j=1}^{N}\left\langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_j}\right\rangle\left\langle \mathrm{e}^{-\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_j}\right\rangle \\ &= \frac{1}{N}\sum_{j=1}^{N}\left|\mathrm{e}^{-\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_j}\right|^2 \qquad (49) \\ &= \frac{1}{N}\sum_{j=1}^{N}\left|\int_V \mathrm{d}^3r\,\exp\left(\mathrm{i}\mathbf{Q}\cdot\mathbf{r}\right)\rho_j(\mathbf{r})\right|^2 . \qquad (50) \end{aligned}$$

Here, $\rho_j(\mathbf{r})$ denotes the 'density of particle j', i.e. the probability density of the individual particle j being at $\mathbf{r}$. One can see that the elastic incoherent structure factor differs from the structure factor (17) itself mainly by the order of summation and averaging. It has the normalisation $S^{\mathrm{el}}_{\mathrm{inc}}(0) = 1$, that of a form factor. One can say that the EISF is the form factor of the volume confining the motion of the particles. E.g. for particles performing any kind of motion inside a sphere, the EISF would be $S^{\mathrm{el}}_{\mathrm{inc}}(Q) = 9\left(\sin(QR) - QR\cos(QR)\right)^2/Q^6R^6$ as given by (21).

As in the static situation, the scattering law can be traced back to distance distribution functions, the *van Hove correlation functions*, which are time-dependent:

$$G(\mathbf{r},t) = \frac{1}{N}\left\langle\sum_{j,k=1}^{N}\delta(\mathbf{r}-\mathbf{r}_k(t)+\mathbf{r}_j(0))\right\rangle , \tag{51}$$

$$G_{\mathrm{s}}(\mathbf{r},t) = \frac{1}{N}\left\langle\sum_{j=1}^{N}\delta(\mathbf{r}-\mathbf{r}_j(t)+\mathbf{r}_j(0))\right\rangle . \tag{52}$$

Insertion into

$$I_{[\mathrm{coh|inc}]} = \int_{V_d} G_{[\mathrm{s}]}(\mathbf{r},t)\exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r})\mathrm{d}^3r \tag{53}$$

directly proves that the spatial Fourier transforms of the van Hove correlation functions are the intermediate scattering functions.

The two particle version can—as in the static case—be reduced to the microscopic density,

$$\rho(\mathbf{r},t) = \sum_{j=1}^{N}\delta(\mathbf{r}-\mathbf{r}_j(t)) . \tag{54}$$

As in the static case $\rho(\mathbf{r})$, $\rho(\mathbf{r},t)$ can also be seen as a generalisation to the situation of scattering from continua. Its autocorrelation function in space and time is

$$\langle\rho(\mathbf{0},0)\rho(\mathbf{r},t)\rangle\,. \tag{55}$$

The $\mathbf{0}$ is again showing that translational symmetry is assumed. So the correlation function can be replaced by its average over all starting points $\mathbf{r}_1$ in the sample volume:

$$\langle\rho(\mathbf{0},0)\rho(\mathbf{r},t)\rangle = \frac{1}{V}\int_V \mathrm{d}^3 r_1 \langle\rho(\mathbf{r}_1,0)\rho(\mathbf{r}_1+\mathbf{r},t)\rangle\,. \tag{56}$$

Insertion of (54) gives

$$\langle\rho(\mathbf{0},0)\rho(\mathbf{r},t)\rangle = \frac{1}{V}\left\langle\sum_{j,k=1}^{N}\int_V \mathrm{d}^3 r_1 \delta(\mathbf{r}_1-\mathbf{r}_k(t))\delta(\mathbf{r}_1+\mathbf{r}-\mathbf{r}_j(t))\right\rangle \tag{57}$$

$$= \frac{1}{V}\left\langle\sum_{j,k=1}^{N}\delta(\mathbf{r}_k(t)+\mathbf{r}-\mathbf{r}_j(t))\right\rangle \tag{58}$$

which with (51) implies

$$G(\mathbf{r},t) = \frac{1}{\rho_0}\langle\rho(\mathbf{0},0)\rho(\mathbf{r},t)\rangle\,. \tag{59}$$

Again setting $t=0$ results in the static scattering situation:

$$G(\mathbf{r},0) = \frac{\langle\rho(\mathbf{0},0)\rho(\mathbf{r},0)\rangle}{\rho_0} = \delta(\mathbf{r}) + \rho_0 g(\mathbf{r}) \tag{60}$$

with $g(\mathbf{r})$ from equation (16).
As in the case of static scattering there is an alternative way to derive the scattering function by first Fourier-transforming the density

$$\rho_{\mathbf{Q}}(t) = \int \mathrm{d}^3 r \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}}\rho(\mathbf{r},t) = \sum_{j=1}^{N}\mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_j(t)} \tag{61}$$

and then multiplying its conjugated value at $t=0$ with that at t:

$$I_{\mathrm{coh}}(\mathbf{Q},t) = \frac{1}{N}\left\langle\rho^*_{\mathbf{Q}}(0)\rho_{\mathbf{Q}}(t)\right\rangle \tag{62}$$

and

$$S_{\mathrm{coh}}(\mathbf{Q},\omega) = \frac{1}{2\pi N}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\omega t}\left\langle\rho^*_{\mathbf{Q}}(0)\rho_{\mathbf{Q}}(t)\right\rangle \mathrm{d}t\,. \tag{63}$$

(This is a general consequence of the *cross-correlation theorem* of Fourier transform (121).)
Note that a reduction of the self correlation function $G_{\mathrm{s}}(\mathbf{r},t)$ is *not* possible in the same way because the multiplication $\rho(\mathbf{0},0)\rho(\mathbf{r},t)$ inevitably includes all combinations of particles j,k and not only the terms for identical particles j,j.
The derivation of inelastic scattering presented here, starting from expression (32), was introduced by Vineyard in 1958 [21]. It was discovered already four years later [22] that this is a very rough approximation, among other shortcomings violating *detailed balance*. While

equations (38) and (39) with reasonable assumptions about the symmetry of the system imply $S(Q,-\omega) = S(Q,\omega)$, detailed balance requires

$$S(Q,-\omega) = \exp\left(\frac{\hbar\omega}{k_\mathrm{B}T}\right) S(Q,\omega)\,. \tag{64}$$

In equilibrium, the probability for the scattering system to be in the lower energy state is higher by the factor $\exp(\hbar\omega/k_\mathrm{B}T)$. Therefore scattering into a state with higher energy is more probable by the same factor than scattering into the lower energy state. This results from thermodynamics and is independent from the approximation to neglect quantum mechanics. Fortunately, in the case of soft matter systems one often deals with slow processes at comparatively high temperatures so that $\hbar\omega \ll k_\mathrm{B}T$ and the approximations used here are valid nevertheless.
A more exact treatment (even on the semi-classical level) would make it necessary to know the momenta of the particles in addition to their trajectories [22]. In case of absence of such information and if it is necessary to access high energy transfers $\hbar\omega$ it is often justified to 'force' detailed balance by an ad hoc factor [23]:

$$S^{\mathrm{corrected}}(Q,\omega) = \exp\left(-\frac{\hbar\omega}{2k_\mathrm{B}T}\right) S(Q,\omega)\,. \tag{65}$$

3.1 Example: diffusion

For simple diffusion the density develops in time following Fick's second law,

$$\frac{\partial\rho}{\partial t} = D\Delta\rho \equiv D\left(\frac{\partial^2\rho}{\partial x^2} + \frac{\partial^2\rho}{\partial y^2} + \frac{\partial^2\rho}{\partial z^2}\right)\,. \tag{66}$$

The underlying mechanism is Brownian motion, i.e. random collisions with solvent molecules. Therefore, it can be concluded from the central limit theorem of statistics that the density of particles initially assembled at the origin is a Gaussian in all coordinates:

$$\begin{aligned}\rho_1 &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y^2}{2\sigma^2}\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{z^2}{2\sigma^2}\right)\\ &= \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\left(-\frac{r^2}{2\sigma^2}\right)\,.\end{aligned} \tag{67}$$

The index 1 should remind that the prefactor is chosen such that the total particle number $\int \rho_1 \ \mathrm{d}^3r$ is normalised to one. The width of the distribution, σ has the dimension length. The only way to construct a length out of D (dimension $\mathrm{length}^2/\mathrm{time}$) and time is $\sigma = c\sqrt{Dt}$ where c is a dimensionless constant. Inserting this into (67) yields:

$$\rho_1 = \frac{1}{c^3(2\pi Dt)^{3/2}}\exp\left(-\frac{r^2}{2c^2Dt}\right)\,. \tag{68}$$

The derivatives of this expression with respect to t and x, y, z can be calculated and inserted into (66):

$$\frac{\sqrt{2}\,(r^2-3c^2Dt)}{8\pi^{3/2}c^5D^{5/2}t^{7/2}}\exp\left(-\frac{r^2}{2c^2Dt}\right) = \frac{\sqrt{2}\,(r^2-3c^2Dt)}{4\pi^{3/2}c^7D^{7/2}t^{7/2}}\exp\left(-\frac{r^2}{2c^2Dt}\right)\,. \tag{69}$$

One can see that the right- and left-hand side are identical if $c = \sqrt{2}$. This proves that the 'guess' (67) is indeed a solution of Fick's second law and also determines the unknown c. With the value of c substituted, the 'single particle density' is

$$\rho_1 = \frac{1}{(8\pi Dt)^{3/2}} \exp\left(-\frac{r^2}{4Dt}\right) . \tag{70}$$

Diffusion-like processes are often characterised by the *mean-square displacement* $\langle r^2 \rangle$. Because of the statistical isotropy, the average displacement $\langle r \rangle$ is always zero. Therefore, the characterisation of the mobility of a diffusional process has to be done using the second moment, which is the average of the square of the displacement. For the simple Fickian diffusion this can be calculated from (70):

$$\langle r^2 \rangle = \int \rho_1 r^2 4\pi r^2 \mathrm{d}^3 r = 6Dt\,. \tag{71}$$

For incoherent scattering the starting position $\mathbf{r}(0)$ is irrelevant. Therefore, expression (70) is also $G_\mathrm{s}(r,t)$. Because the Fourier transform of a Gaussian function is a Gaussian itself, the corresponding incoherent intermediate scattering function is

$$I_\mathrm{inc}(Q,t) = \exp\left(-DQ^2 t\right) , \tag{72}$$

and because the Fourier transform of an exponential decay is a Lorentzian the incoherent scattering function is

$$S_\mathrm{inc}(Q,\omega) = \frac{1}{\pi} \frac{DQ^2}{\omega^2 + (DQ^2)^2} . \tag{73}$$

This function is centred around $\omega = 0$, and for that reason the scattering is called *quasielastic*. This is typical for diffusionlike processes in contrast to vibrational processes which yield (phonon) peaks at finite energy transfers. For this reason, many textbook authors distinguish between inelastic and quasielastic neutron scattering instead of subsuming the latter under the former as done here.

From expression (72) one can see that $I_\mathrm{inc}(Q,t)$ decays faster with time for larger Q and from (73) that $S_\mathrm{inc}(Q,\omega)$ is getting broader. This is understandable because Q defines the spatial resolution of a neutron scattering experiment in a reciprocal way. So a larger Q means observation on shorter distances which can be travelled faster by the diffusing particle.

Finally, one can see that

$$I_\mathrm{inc}(Q,t) = \exp\left(-\frac{Q^2\langle r^2 \rangle}{6}\right) . \tag{74}$$

Because this expression is derived independently of the specific form of $\sigma(t)$ in (67) it is generally valid if the distribution of displacements $G_\mathrm{s}(r,t)$ is a Gaussian. Even if this is not the case, equation (74) is often a good low-Q approximation called the *Gaussian approximation* and is the dynamical analogue of to the Guinier approximation of static scattering.

In general, the incoherent intermediate scattering function cannot be derived from the mean-square displacement alone. Because equation (74) is the first term of the cumulant expansion $\exp(aQ^2 + bQ^4 + \dots)$ of $I_\mathrm{inc}(Q,t)$ [24] the mean-square displacement can be calculated as

$$\langle r^2 \rangle = -\lim_{Q\to 0} \frac{6}{Q^2} \ln I_\mathrm{inc}(Q,t) \quad \text{or} \tag{75}$$

$$\langle r^2 \rangle = -\left.\frac{\mathrm{d} \ln I_\mathrm{inc}(Q,t)}{\mathrm{d}Q^2}\right|_{Q=0} . \tag{76}$$

By replacing $I_{\mathrm{inc}}(Q,t)$ by its value at infinite time, the EISF $S^{\mathrm{el}}_{\mathrm{inc}}(Q)$, the limiting mean-square displacement of a confined motion can be obtained. This is the principle of the *elastic scan* technique often used on neutron backscattering spectrometers [25].

4 Scattering probes: photons, neutrons, electrons

The most obvious difference between different scattering probes, i.e. radiation types, is their wavelength and frequency. In the particle picture these correspond to the momentum ($p = h/\lambda$) and the energy ($E = hf$) of the particles. For each radiation type there is a relation between both quantities based on the rest mass of the particle, m_0:

$$\lambda = \frac{h}{\sqrt{2Em_0 + E^2/c^2}}\,. \tag{77}$$

Note that E is the net kinetic energy here, thus $E = (m - m_0)\,c^2$ relativistically. For the photon, a massless particle, the relation reduces to

$$\lambda = \frac{hc}{E} = \frac{c}{f} \tag{78}$$

with $f = E/h$ being the (common) frequency. So for photons the wavelength decreases inverse-proportionally with frequency, $\lambda \propto E^{-1}$. Conversely in the non-relativistic limit ($m_0 \gg E/c^2$) which would be fulfilled for neutrons in general, one obtains:

$$\lambda = \frac{h}{\sqrt{2Em_0}} = \frac{h}{p} \tag{79}$$

with p being the momentum of the particle. Thus for massive (non-relativistical) particles the wavelength only decreases as $\lambda \propto E^{-1/2}$.

Fig. 6 shows the relation between wavelength and energy/frequency for photons, neutrons, and electrons. It can be seen that in the determination of molecular structure, neutrons and x-rays are equivalent because they both have adequate wavelengths. Nevertheless for inelastic scattering experiments x-rays have an energy several orders of magnitude higher at the same wavelength. Therefore the energy transfers which correspond to processes in a time range of nanoseconds and above are much more difficult to resolve by x-ray scattering.

In order to dwell on this argument in more detail, in the following it is explained how the wavelength and energy of the probe beam determine the resolution and range of a scattering method in time and space:

- The *minimal distance* resolved in space is related directly to the wavelength. This can be seen from Bragg's law, $\lambda = 2d\sin\theta$. Because $\sin\theta \leq 1$ the distance (of Bragg planes) is limited by $d \geq \lambda/2$ [6]. Expressed in terms of the scattering vector, equation (4), this corresponds to an upper limit of the scattering vector $Q \leq 4\pi/\lambda$.

[6] It has to be noted that in the absence of a crystalline structure this relation is only an estimate. In principle, the whole range of distances r enters a Fourier transform expression as (18) for a given Q. Nevertheless, a relation as $d_{\min} = 2\pi/Q_{\max}$ gives the smallest structure which can be technically resolved also in this case, but with a prefactor which may differ from 2π depending on the definition of what "resolve" means.

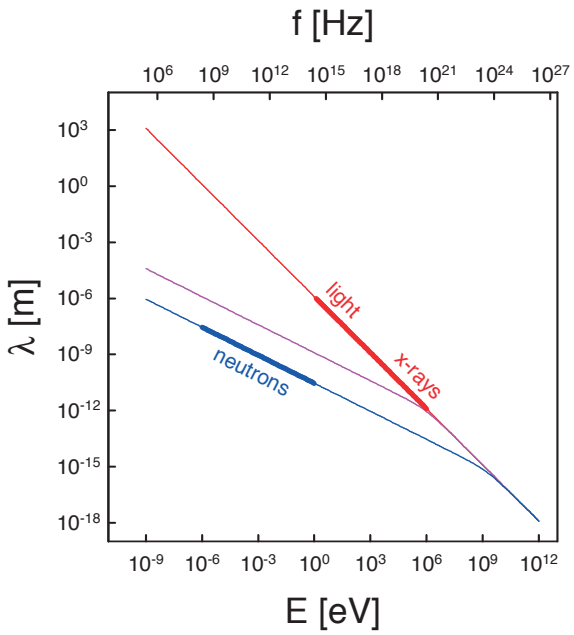


Fig. 6: *Relation between energy (E) / frequency (f) and wavelength (λ) of different scattering probes. Red: photons, pink: electrons, blue: neutrons. The broadened sections of the curves represent the ranges actually used in scattering experiments on soft matter systems.*

- The *maximal distance* accessible is also proportional to the wavelength, but here the instrumental resolution is also a determining factor. Again in the picture of Bragg's law, if the smallest technically accessible scattering angle is $\theta_{\min}$ then the largest observable Bragg plane distance is $d \leq \lambda/2 \sin\theta_{\min}$. So here the wavelength of the beam does not pose a limit on principle. Nevertheless, even on dedicated small-angle scattering instruments one is not able to reach angles below $0.01°$. Therefore, there is a technological limit at about four orders of magnitude above the wavelength of the radiation used in a scattering experiment, $d \lesssim 10^4\lambda$.

- The *minimal time* resolvable in dynamics is inversely related to the frequency of the radiation, or the particle energy. Here, the argument is somewhat more complicated because the energy *transfer* $E = \hbar\omega$, especially in neutron scattering, may be *larger* than the energy hf. But as Fig. 5 shows, large energy transfers lead to a strongly offset Q range. So there is a combined restriction of time- and space range which usually prevents energy transfers to be used which are more than an order of magnitude higher than the incident energy. This leads to a restriction $t > 0.2\,\mathrm{ps\ meV}/E$.

- For the *maximal time* accessible, as in the case of the maximal distance, the instrumental resolution is a crucial factor. The limit in terms of the minimal resolvable energy transfer is $t < 2\,\mathrm{ps\ meV}/\Delta E$ (with a slight dependence of the prefactor on the definition of "accessible time"). But the relative instrumental resolution $\Delta E/E$ differs by orders of magnitude from instrument to instrument. Conventional neutron scattering instruments reach down to 10^{-3} while due to the high flux available inelastic experiments on syn-

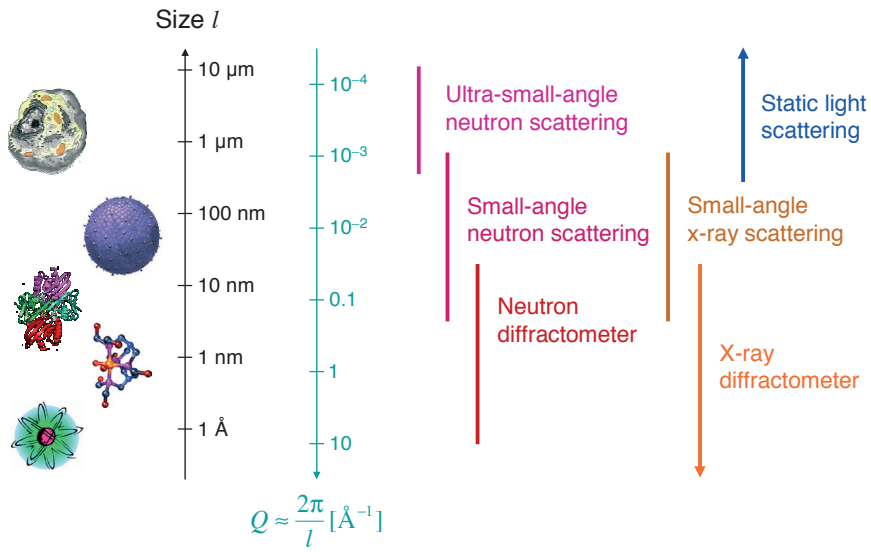


Fig. 7: *Spatial dimensions accessible by different scattering methods. (Note that these are typical ranges; particular experimental set-ups may extend these.)*

chrotrons can reach 10^{-7}. The neutron spin echo technique (see subsection 5.3) operating directly in the time domain can reach times corresponding to $\Delta E/E = 10^{-6}$. A method opening an on-principle unlimited time range is the use of *intensity autocorrelation* on the scattered beam. For this method the resolution is not restricted by the monochromatisation of the incident beam. This is a standard technique for photons, where times up to 10^3 s can be reached. The corresponding technique for x-rays is currently under development reaching roughly the same limit in certain applications.

Fig. 7 shows the spatial dimensions accessible by different elastic scattering methods. It can be seen that for neutrons and x-rays usually small scattering angles have to be used to reach the dimensions of soft matter. On the other hand, light scattering is only suited to very large soft matter objects.

Fig. 8 shows in dimensions of space and time the range of different inelastic scattering methods. It can be seen that light scattering covers an enormous range in the dynamics due to the availability of the photon correlation technique. On the other hand it suffers from the long wavelength prohibiting access to shorter length scales. X-rays do cover nanometre sizes but due to their high energy only allow to resolve very fast processes ($\lesssim 1$ ps) with filter methods. X-ray photon correlation methods are currently limited to longer times than some microseconds. There remains a large region, roughly the nanometre-nanosecond range, which can only be covered by inelastic neutron scattering.

Another property of a probe beam important for the application is its interaction mechanism with the sample. Photons of long wavelength (light) interact with the dielectric function of the sample. Thus, any sample showing a fluctuation of the index of refraction is suitable for scattering. On the other hand, short-wavelength photons, x-rays, interact directly with the electron density. Therefore, their interaction is stronger with elements of higher atomic number Z (Fig. 9). At roughly the same wavelength neutrons do not show the monotonic dependence on Z and in addition usually the scattering length depends on the isotope. Especially, the scattering

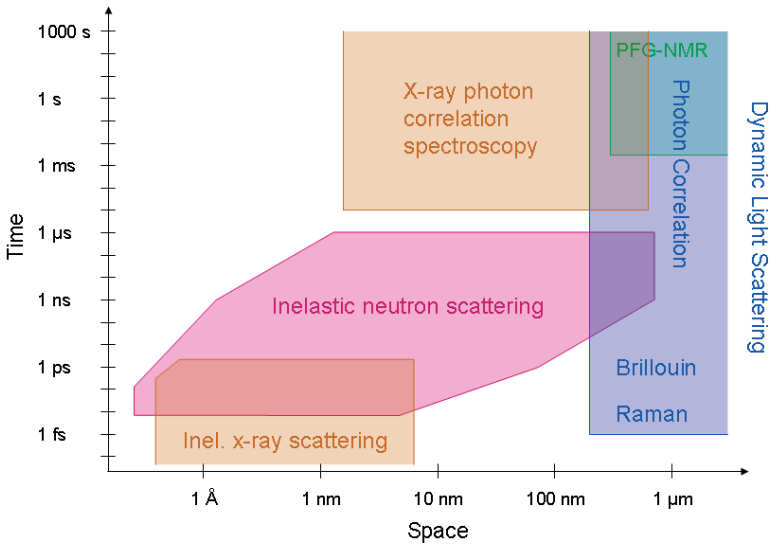


Fig. 8: *Ranges in time and space accessible by different scattering methods. PFG-NMR has been added because it provides the intermediate scattering function $I(Q,t)$ as genuine scattering methods do. (Note that these are typical ranges; particular experimental set-ups may extend these.)*

lengths of hydrogen and deuterium are completely different. In organic materials this allows to create a contrast without change of the chemical properties.

Finally, neutron scattering differs from photon scattering by that due to their magnetic moment neutrons may also be scattered by magnetic interactions. The direct application of this feature is of less relevance in soft matter science because few such materials contain magnetic elements. Nevertheless, the incoherent scattering discussed in section 3 is often (especially in the case of hydrogen) due to spin-disorder. Because spin-incoherent scattering flips the spin of the scattered neutron with a probability of $2/3$ the total scattering can be separated into coherent and incoherent using the following formulae:

$$\frac{\partial\sigma_{\rm coh}}{\partial\Omega\partial\omega} = N\frac{k'}{k}\frac{\sigma_{\rm coh}}{4\pi}S_{\rm coh}(\mathbf{Q},\omega) = \frac{\partial\sigma_{\uparrow}}{\partial\Omega\partial\omega} - \frac{1}{2}\frac{\partial\sigma_{\downarrow}}{\partial\Omega\partial\omega}\,, \tag{80}$$

$$\frac{\partial\sigma_{\rm inc}}{\partial\Omega\partial\omega} = N\frac{k'}{k}\frac{\sigma_{\rm inc}}{4\pi}S_{\rm inc}(\mathbf{Q},\omega) = \frac{3}{2}\frac{\partial\sigma_{\downarrow}}{\partial\Omega\partial\omega}\,. \tag{81}$$

Here $\partial\sigma_{\uparrow}/\partial\Omega\partial\omega$ and $\partial\sigma_{\downarrow}/\partial\Omega\partial\omega$ denote the double-differential cross-section separated by polarisation analysis into a part without and with spin flip respectively. For inelastic neutron scattering this allows a distinction between the self- and collective part of the dynamics. For diffraction experiments, property (44) provides a possibility to calibrate the intensity on the incoherent scattering an internal reference. This method is often more reliable than calibration by an external standard (e.g. vanadium).

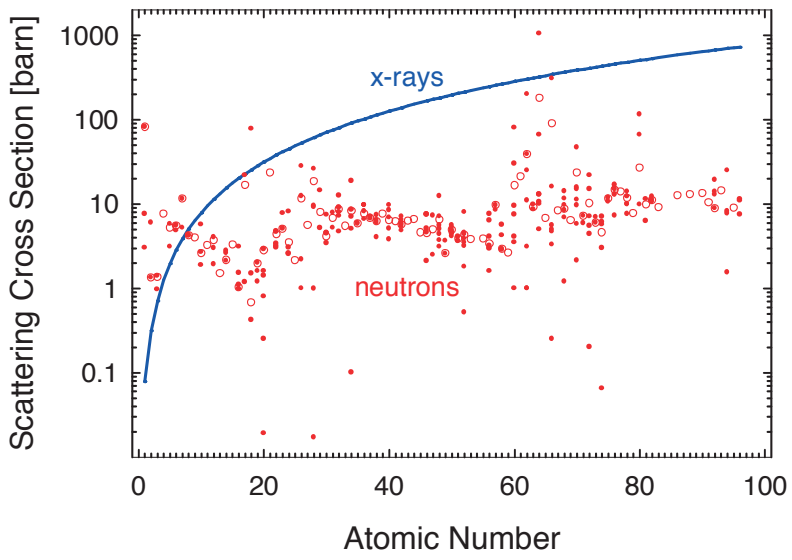


Fig. 9: *Scattering cross sections for neutrons and x-rays. Red symbols: neutron scattering cross section* $\sigma_{\rm coh} + \sigma_{\rm inc}$. *The small points indicate individual isotopes and the large open circles averages for the natural composition of the elements. Blue line: x-ray scattering cross section. The unit of the cross section is* $1\,\mathrm{barn} = 10^{-28}\,\mathrm{m}^2$.

5 Scattering instruments for soft matter investigations

5.1 Light scattering

Static light scattering is deployed to investigate the structure of soft matter particles (polymers, colloids, proteins…) [8]. In order to get information about the detailed structure in terms of a scattering length density or correlation function it is necessary that this structure is larger than about half the wavelength[7]. Therefore, static light scattering is restricted to rather large objects. Nevertheless, if the objective is just a *characterisation* of particles with respect to their molecular weight it is also possible to obtain this from the low-Q range. For this purpose the fact that the scattering only depends on the squared number of scatterers for low Q is used. Let us assume $\mathcal{N}$ species of particles for each of which there are $N_{\mathrm{p}i}$ particles with $N_{\mathrm{s}i}$ scatterers. Then the scattering from one particle is from (3): $\mathrm{d}\sigma/\mathrm{d}\Omega|_{Q\to 0} = b^2 {N_{\mathrm{s}i}}^2$. If one assumes that there is no correlation between the particles, the scattering normalised to the total number of scatterers is

$$\frac{b^2 \sum_{i=1}^{\mathcal{N}} {N_{\mathrm{s}i}}^2}{\sum_{i=1}^{\mathcal{N}} N_{\mathrm{s}i}} = \frac{b^2 \overline{{N_{\mathrm{s}i}}^2}}{\overline{N_{\mathrm{s}i}}} \,. \tag{82}$$

If the particles are *chemically* identical their mass and molecular weight is proportional to the number of scatterers. Thus, the left hand side is proportional to the scattering divided by the

[7] This restriction is valid if the strict sense of "structure" as full knowledge of the molecular shape is applied as defined in footnote 1. The determination of certain properties as molecular weight and radius of gyration are possible also for smaller molecules. The limit in this sense would be about 10 nm for static light scattering and 1 nm for dynamic.

concentration and the right hand side proportional to the *weight-averaged* molecular weight:

$$\frac{1}{c}\left.\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right|_{Q\to 0} \propto \overline{M_\mathrm{w}}\,. \tag{83}$$

The proportionality factor can be calculated for light scattering from the index of refraction and its derivative with respect to the concentration [26]. In the terms of equation (31) this result relies on $P(Q) = 1$ requiring $Q \to 0$ and $S(Q) = 1$ (no interparticle interaction) requiring $c \to 0$ [8]. The two limits are often graphically realised in the so-called Zimm plot [26,27]. From such a plot, as a measure of deviation from the $c \to 0$ limit, the second virial coefficient can be extracted too.

The by far most important *dynamic* light scattering (DLS) technique used for soft matter systems is photon correlation spectroscopy. Only a short introduction will be given here leaving the details to standard textbooks [9–12,28]: Monochromatic laser light of wavelength λ is scattered from the sample. The scattered photons are registered in a detector giving a time dependent intensity signal $I(t)$. From this signal a special purpose computer ("correlator") constructs the correlation function

$$\langle I(0)I(\tau)\rangle = \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} I(t)I(t+\tau)\mathrm{d}t\,. \tag{84}$$

(Of course $I(t)$ and also the correlation functions depend on the scattering angle and therefore on $\mathbf{Q}$. To simplify the expressions this dependence will only be shown where necessary.) The practical meaning of this equation is the following: The integral with the factor $1/T$ is the average over the time interval of length T. In order to obtain a stable value independent of the short-time fluctuations one has to take the limit of infinite time interval length T. For comparison the average intensity is defined by

$$\langle I\rangle = \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} I(t)\mathrm{d}t\,. \tag{85}$$

In contrast to this definition equation (84) contains the product of two intensity measurements which are apart by the time τ. Therefore, $\langle I(0)I(\tau)\rangle$ expresses in how far intensities measured in this temporal distance are correlated. It is easy to see that if there is no temporal correlation (84) reduces to the square of (85), $\langle I\rangle^2$. On the other hand if there is complete correlation (especially for $\tau = 0$) it assumes the value $\langle I^2\rangle$ which is always greater or equal to the former value.

In order to express the degree of correlation independently of the intensity itself a normalised quantity is defined

$$g_2(t) = \frac{\langle I(0)I(t)\rangle}{\langle I\rangle^2}\,. \tag{86}$$

Under certain ideal experimental conditions this function decays from 2 to 1. In order to obtain a connection to the microscopic properties of a soft matter system one needs the correlation function of the electric field instead of that of the intensities, namely

$$g_1(t) = \frac{\langle E(0)E(t)\rangle}{\langle E^2\rangle}\,. \tag{87}$$

[8] For light scattering there is also an additional Q-dependent term, $1 + \cos^2 2\theta$ for unpolarised light, resulting from the dependence of the cross section on the mutual orientation of the polarisation of incident and scattered beam.

With certain assumptions, the most important being that the electrical field is a Gaussian-distributed quantity, one can derive the *Siegert relation* between the two correlation functions:

$$g_1(t) = \sqrt{g_2(t) - 1}\,. \tag{88}$$

Light scattering theory gives a simple relation between the field correlation function (87) and the time dependent particle positions of a liquid-like system assuming its constituting particles are identical (monodisperse) with respect to their shape, size, and optical properties:

$$g_1(t) = \frac{I(\mathbf{Q}, t)}{S(\mathbf{Q})}\,. \tag{89}$$

Here, $S(\mathbf{Q})$ is the static structure factor (17) and $I(\mathbf{Q}, t)$ the dynamic structure factor or intermediate (in most cases: coherent) scattering function (41). Because of the larger wavelength of light, these quantities refer here to the position of particles (e.g. polymer molecules) and not to individual atoms as in x-ray or neutron scattering. Therefore, the $S(Q)$ here is rather the one in expression (31) but since the form factor $P(Q)$ would show up in the numerator and denominator of (89) equally it cancels out. This implies a drawback of a typical DLS experiment: Only the centre-of-mass motion of a particle is observed, not the internal dynamics. Observation of the latter is only possible for very large objects (with typical dimensions of the order of the wavelength of light) or by inelastic neutron scattering due to the neutrons' shorter wavelength.
DLS can also be used as a particle sizing method, namely to obtain information about the molecular weight distribution beyond a single moment as (83). The principle of this application is that for polydisperse systems equation (72) changes into

$$I(Q, t) = \int_0^\infty p(D) \exp(-DQ^2 t) \mathrm{d}D \tag{90}$$

where $p(D)$ is the distribution of diffusion coefficients normalised as $\int_0^\infty p(D)\mathrm{d}D = 1$. The inversion of (90) yielding the distribution $p(D)$ from the experimentally obtained $I(Q, t)$ is a mathematically *ill-posed problem*. Therefore one needs special regularisation algorithms as the widely used CONTIN [29] for this purpose or has to restrict oneself to obtaining several moments of the distribution by the *method of cumulants* [30]. The thus obtained distribution of diffusion coefficients has then to be converted into a distribution of particle sizes by the Stokes-Einstein relation

$$D = \frac{k_\mathrm{B} T}{6\pi R_\mathrm{H} \eta} \tag{91}$$

where R_H is the (hydrodynamic) radius of the particle and η the viscosity of the solvent.
In Raman- and Brillouin scattering inelastic light scattering is observed by filter methods yielding $S(Q, \omega)$ in the frequency domain. These methods are not such important in the soft matter studies because their energy resolution corresponds to rather high frequency which at the Q values given for light scattering would only allow to observe very fast processes. Only for the observation of glassy dynamics in molecular liquids and polymers especially the Fabry-Pérot interferometer has some relevance [31].

5.2 X-ray scattering

Because soft matter object dimensions ($\gtrsim 1\,\mathrm{nm}$) usually exceed the wavelength of x-rays ($\lesssim 0.1\,\mathrm{nm}$) by more than an order of magnitude, x-ray scattering experiments require small angles.

Small-angle x-ray scattering (SAXS) is conceptually not different from ordinary diffraction but the small scattering angle often requires some technical 'tricks': (1) All such instruments have to employ an evacuated flight path to avoid scattering from air. (2) The small solid angles of collimation and detection lower the intensities. Therefore either strong sources are necessary (rotating anode tubes for tabletop applications, or a synchrotron) or point collimation has be replaced by slit collimation (Kratky camera [32]). In the latter case the gain in intensity has to be paid for by a distortion of the scattering curves which has to be corrected mathematically. Several more advanced SAXS set-ups are discussed in ref. 33. (3) Because soft matter usually consists of elements with lower atomic number than their sample containers, extremely thin-walled containers have to be used (e.g. Mark capillaries).

Inelastic x-ray scattering methods can be divided analogously to light scattering into those operating by filters in the energy-frequency domain and by photon correlation in the time domain. In both cases, due to the required resolution the efficiency is low. Therefore, only synchrotrons come into consideration as sources.

Typical current instruments using the first approach, as the set-up for inelastic x-ray scattering on beamline ID16 of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France [34], use incident photon energies of $5 \ldots 20\,\mathrm{keV}$. This corresponds to wavelengths of $0.06 \ldots 0.25\,\mathrm{nm}$. This is about the same range as for SAXS and thus would allow observing typical soft matter dimensions in space. The energy resolution achievable with these incident energies is $1 \ldots 10\,\mathrm{meV}$. This is a remarkable experimental feat, corresponding to $\Delta E/E \approx 10^{-7}$. Nevertheless, the corresponding times are $0.2 \ldots 2\,\mathrm{ps}$, clearly too short for observing most processes in soft matter systems.

The more promising approach is the use of photon correlation [35], as on beamline ID10A at the ESRF [36], x-ray photon correlation spectroscopy (XPCS). There, a moderately coherent beam is selected from an undulator spectrum. This is done by crystal monochromatisation ($\Delta E/E \approx 10^{-4}$) and an extreme collimation of the beam ($\approx 10\,\mu\mathrm{m}$ sized apertures on flight paths of several metres) The subsequent instrument is similar to the set-up used in light PCS. As detector either point detectors with a low 'dead time' (down to a nanosecond for avalanche photodiodes) or CCD cameras which allow the simultaneous registration of a 2-dimensional **Q** range are used. The drawback of the latter is that the readout time limits the shortest accessible time to about $20\,\mathrm{ns}$. The incident wavelengths used are similar to those of the previous x-ray techniques. Therefore, concerning the spatial resolution the limitation of light PCS towards smaller structures can be overcome. On principle, a similarly large time range can be expected as in the case of light PCS. (Of course, also here $g^{(2)}(t)$ is measured and not $I(Q,t)$.) Nevertheless, there are several technical problems which restrict the efficiency currently: (1) Due to the extremely restrictive selection, even for synchrotrons, the count statistics is low (especially if compared to what is available by lasers). Therefore, the short time limit is often not given by the instrumental limitation of the detector-correlator unit (which would be at about $10^{-8}\,\mathrm{s}$ for point detectors) but the fading statistics. For that reason the first successful XPCS experiments were all on high x-ray contrast materials as e.g. metal colloids. Additional noise at short times may arise from particular 'filling modes' of the synchrotron beam. (2) Nevertheless, the flux at the sample is often so high ($66\,\mathrm{W/mm^2}$ for a typical set-up of ID10A) that beam damage is the consequence for soft matter samples. (3) The long end of the time range is limited indirectly by the lifetime of the beam. Because the averaging time of the correlation function has to be about $1/100$ of the lifetime, the effective limit for currently realised lifetimes of $10 \ldots 80\,\mathrm{h}$ of synchrotron beams is around $1000\,\mathrm{s}$.

5.3 Neutron scattering

Neutron scattering experiments on soft matter typically use 'cold neutrons', i.e. neutrons where the kinetic energy is reduced by multiple collisions in a moderator which is kept far below room temperature. A typical temperature would be $T = 20\,\mathrm{K}$ corresponding to an energy $k_\mathrm{B}T = 1.7\,\mathrm{meV}$ and a wavelength $\lambda = 0.7\,\mathrm{nm}$. Although this wavelength is closer to, e.g., the dimensions of a polymer molecule than that of thermal neutrons ($\approx 0.2\,\mathrm{nm}$), the observation of its chain structure usually requires small scattering angles (SANS) [33, 37]. The standard instrument is a scaled-up pinhole SAXS set-up. Because the width of the neutron beam is usually about a centimetre, 10–100 times larger than an x-ray beam, the whole instrument has to be enlarged by this factor resulting in flight paths up to 40 m. With this conventional instrument, scattering vectors down to $Q = 10^{-2}\,\mathrm{nm}^{-1}$ can be achieved. For lower Q values, focussing elements as mirrors or neutron lenses have to be implemented. The beam size is reduced by apertures to some millimetres and the focussing partially compensates the loss in intensity. The limit of this technology lies at about $10^{-3}\,\mathrm{nm}^{-1}$. For even lower Q, Bonse-Hart set-ups are used as in x-ray scattering (limit $\approx 10^{-4}\,\mathrm{nm}^{-1}$).

The crucial advantage of SANS compared to SAXS, justifying the far higher experimental effort, is the possibility of *contrast variation*. While in x-ray scattering the cross-section is invariably coupled to the atomic number, it can be varied in neutron scattering by isotopic substitution. This allows e.g. to create a scattering contrast between chemically identical molecules. This 'trick' was used for the first unambiguous confirmation of the Gaussian-coil structure of a polymer in a melt [38]. Conversely, one can make structures 'invisible' by 'contrast-matching' certain microphases in a sample. This is done by choosing an isotopic composition which makes their average scattering length densities (8) identical.

The natural time scale of neutrons $\hbar/E$ is rather short in comparison to the times relevant for soft matter dynamics. Even for cold neutrons it lies at about one picosecond. Therefore, inelastic neutron scattering experiments need a very high resolution $\Delta E/E$. There are currently three techniques used for soft matter [39]—time-of-flight spectroscopy, backscattering spectroscopy, and neutron spin-echo—of which the last is arguably the most important one.

Neutron scattering time-of-flight (TOF) spectroscopy [40] uses the fact that connected to their energy the velocity of neutrons is changed when they are scattered inelastically. In the instrument a pulsed monochromatic ($E = \mathrm{const.}$) neutron beam impinges on the sample. After being scattered, the neutrons reach the detector passing a flight path of some metres. From the time between the release of the pulse and the detector signal the velocity can be calculated and from this the energy after scattering, E'. Thus $\hbar\omega = E' - E$ can be calculated and a spectrum of energy transfers can be constructed. This is done using a multidetector which covers a large solid angle. In this way a Q-ω area as shown in Fig. 5 can be recorded simultaneously. This makes the TOF spectrometer a highly efficient inelastic neutron scattering instrument. The realistic limit of the resolution for TOF is $\Delta E = 10\,\mu\mathrm{eV}$. This results from the problem that a further sharpening of the monochromatisation would lead to an intolerable reduction of intensity. The corresponding time scale is about $200\,\mathrm{ps}$ which is still rather short compared to that of soft matter dynamics.

The backscattering spectrometer (BSS) is a special development of neutron scattering crystal spectrometers optimised for high energy resolution [41]. Crystal spectrometers (for inelastic neutron as well as x-ray scattering) use one or more crystals to monochromatise the beam before scattering by Bragg reflection and detect a single energy after scattering in the same way. Noting that the Bragg relation gives $\Delta\lambda = \cot\theta\Delta\theta$ for the width of the wavelength distribu-

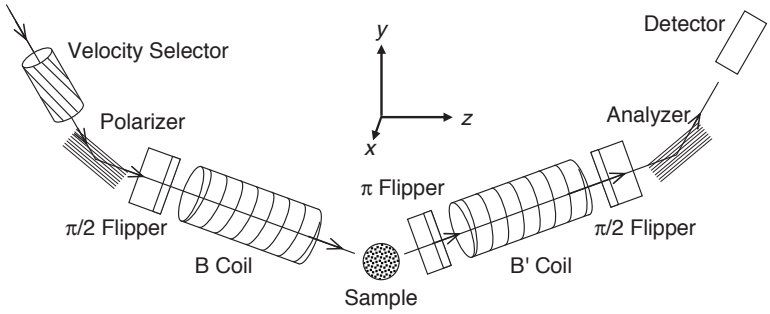


Fig. 10: *Schematic setup of a neutron spin echo spectrometer. x, y, and z denote the general orientation of the coordinate system used in the text with the reservation that z should follow the neutron flight direction.*

tion resulting from an angular error $\Delta\theta$, one notices that the optimally selective position is at $2\theta = 180°$, i.e. under backscattering conditions. (The intricate construction of the instrument is described in the literature [39, 41].) In this way a resolution $\Delta E = 0.3\,\mu\mathrm{eV}$ can be achieved corresponding to about $6\,\mathrm{ns}$.

For even longer times it is not possible anymore to use instruments which define E and E' with sufficient accuracy due to the loss of intensity. Then one needs a method which directly accesses $E' - E$. Such a technique is provided by neutron spin echo (NSE) spectrometers [40, 42, 43]. Because of its importance in the study of soft matter dynamics [44] it will be described in more detail now.

Unlike the previous methods, NSE measures the *individual* velocities of the incident and scattered neutrons using the Larmor precession of the neutron spin in a magnetic field. The neutron spin vector acts as the hand of an internal clock which is linked to each neutron and connects the result of the velocity measurement to the neutron itself. Thereby the velocities before and after scattering on one and the same neutron can be compared and a direct measurement of the velocity difference becomes possible. The energy resolution is thus decoupled from the monochromatisation of the incident beam. Relative energy resolutions in the order of 10^{-5} can be achieved with an incident neutron spectrum of 20% bandwidth.

The motion of the neutron polarisation $\mathbf{P}(t)$—which is the quantum mechanical expectancy value of the neutron spin—is described by the Bloch equation

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \frac{\gamma\mu}{\hbar}(\mathbf{P} \times \mathbf{B}) \tag{92}$$

where γ is the gyromagnetic ratio ($\gamma = -3.82$) of the neutron, μ the nuclear magneton and $\mathbf{B}$ the magnetic field. Equation (92) is the basis for manipulation of the neutron polarisation by external fields. A simple calculation shows that (92) predicts a precession with the *Larmor frequency* $\gamma\mu B/\hbar$ in a constant magnetic field. Thus, if a neutron of wavelength λ is exposed to a magnetic field B over a length l of its flight path its spin is rotated by

$$\phi = \left(\frac{2\pi|\gamma|\mu\lambda m_\mathrm{n}}{h^2}\right) Bl\,. \tag{93}$$

The basic setup of an NSE spectrometer is shown in figure 10. A velocity selector in the primary neutron beam selects a wavelength interval of 10–20% width. In the primary and

Table 1: *Evolution of the neutron spin (classical) in the NSE spectrometer. The left column contains the cartesian components of the normalised spin vector* **s***. The right column shows the neutronic devices the beam passes on its way through the spectrometer. z is defined as always denoting the direction parallel to neutron propagation.*

(s_x, s_y, s_z)	neutronic device
(0,0,1)	
	$\pi/2$ flipper
(1,0,0)	
	field B
$(\cos\phi, \sin\phi, 0)$	
	π flipper
$(\cos\phi, -\sin\phi, 0) =$ $(\cos(-\phi), \sin(-\phi), 0)$	
	field B'
$(\cos(\phi' - \phi), \sin(\phi' - \phi), 0)$	
	$\pi/2$ flipper
$(0, \sin(\phi - \phi'), \cos(\phi - \phi'))$	

secondary flight path of the instrument precession fields B and B' parallel to the respective path are generated by cylindrical coils. Before entering the first flight path the neutron beam is polarised in forward direction[9]. Firstly, a $\pi/2$ flipper rotates the polarisation to the x direction perpendicular to the direction of propagation (z). This is done by exposing the neutrons to a well-defined field for a time defined by their speed and the thickness of a flat coil (Mezei coil). Beginning with this comparatively well-defined initial condition the neutrons start their precession in the field B. After being scattered by the sample the neutrons go through a π flipper. The effect of the π flipper amounts to reverting the precession angle accumulated before scattering. Then the neutrons pass through the second precession field B'. Finally, the neutrons pass through another $\pi/2$ coil which under certain conditions restores their initial polarisation parallel to their flight direction. In order to understand what these conditions are one has to trace the changes of the spin vector as shown in Table 1. In total, the spin is rotated by $\phi - \phi'$ around the x axis when a neutron passes through the spectrometer. This means that the final polarisation is identical to the incident if $\phi = \phi'$ $(+2\pi n)$, especially if $\lambda_\mathrm{i} = \lambda_\mathrm{f}$ (elastic scattering) and $\int_0^l B\mathrm{d}z = \int_0^{l'} B'\mathrm{d}z$ (for homogeneous fields: $Bl = B'l'$) as follows from (93). This condition is called "spin echo" and is independent of the individual velocities of the neutrons because their difference alone determines $\phi - \phi'$.

Leaving spin echo condition, the probability of a single neutron to reach the detector is reduced due to the polarisation analyser by $\cos(\phi' - \phi)$. If we keep the symmetry of the instrument, $Bl = B'l'$, but consider inelastic scattering the precession angle mismatch follows from eq. (93):

$$\begin{aligned} \phi' - \phi &= \left(\frac{2\pi|\gamma|\mu m_\mathrm{n}}{h^2}\right) Bl(\lambda_\mathrm{f} - \lambda_\mathrm{i}) \\ &\approx \underbrace{\frac{|\gamma|\mu {m_\mathrm{n}}^2 \lambda^3 Bl}{h^3}}_{=t_\mathrm{NSE}(B)} \omega \end{aligned} \tag{94}$$

[9]This is done by a 'polarising supermirror' [45] which only reflects neutrons of that spin—similar to the Nicol prism in optics.

The approximation in the second line is valid for small energy transfers where $\Delta\lambda \approx \hbar\omega \,/\, \frac{\mathrm{d}E}{\mathrm{d}\lambda}$ can be used. Because the energy transfer for inelastic scattering is not fixed but distributed as determined by the scattering function $S(\mathbf{Q},\omega)$ we have to average the factor $\cos(\phi' - \phi)$ weighted by $S(\mathbf{Q},\omega)$ to get the reduction of count rate at the detector, the effective polarisation

$$P(\mathbf{Q}, t_{\mathrm{NSE}}) = \frac{\int_{-\infty}^{\infty} S(\mathbf{Q},\omega)\cos(\omega t_{\mathrm{NSE}})\mathrm{d}\omega}{\int_{-\infty}^{\infty} S(\mathbf{Q},\omega)\mathrm{d}\omega}\,. \tag{95}$$

Firstly, we note that $S(\mathbf{Q},\omega)$ in this expression usually is the coherent scattering function. In principle, similar arguments can be used for incoherent scattering because a well-defined fraction of neutrons changes its spin. This leads to a "negative echo" because the majority of neutrons invert their polarisation. But because this effect is only partial (e.g. 2/3 for hydrogen nuclei) it is much more difficult to observe. Only recently, NSE spectroscopy could be applied successfully to incoherently scattering samples [46].
Secondly, expression (95) reverses the temporal Fourier transform[10] in equation (38) or (39). Therefore, NSE provides the intermediate scattering function normalised to the structure factor

$$P(\mathbf{Q}, t_{\mathrm{NSE}}) = \frac{I(\mathbf{Q}, t_{\mathrm{NSE}})}{S(\mathbf{Q})} \tag{96}$$

in the coherent case or just the intermediate scattering function in the incoherent case. Note, that this is formally the same result as for dynamic light scattering (DLS), equation (89). Nevertheless, there are two differences: 1. $I(Q,t)$ and $S(Q)$ here are derived from atomic or molecular positions because of the larger Q used. 2. There is no intermediate step of an intensity correlation involved, removing the uncertainties related to application of the Siegert relation in DLS.
From equation (94) can be seen that the maximum time accessible by NSE increases with the third power of the incident neutron wavelength λ. Therefore, a statement about its limits depends strongly on the required Q value because large Q makes a small λ necessary (eq. (4)). Also for large λ there may be intensity problems because long-wavelength neutrons are not sufficiently present in the source spectrum. For standard applications on polymer chain dynamics the limit in Fourier time lies at about 350 ns.
Many variations of the basic NSE concept explained here are used in currently operated NSE spectrometers. The most straightforward, employed at most such instruments, is the use of multidetectors to observe several Q values simultaneously. The technical problem resulting from this measure is that the field integral has to be kept identical for all possible neutron flight paths by elaborate correction coils [47]. For NSE instruments which are mostly operated at low Q, similar focussing techniques as in (static) small-angle scattering can be used, e.g. mirrors [48]. NSE can be combined with the TOF principle in the primary spectrometer. The components of a neutron pulse enter the spectrometer in the sequence from fast to slow. Thus, the NSE spectrometer can be operated sequentially with all incident wavelengths contained in the pulse. In this way a broad range of Q-t combinations is covered simultaneously. For pulsed sources this leads to a clear intensity gain [49]. It is also possible to use the TOF-NSE principle at a continuous source [48], but then the advantage is partially outweighed by the loss of neutrons during the shaping of the pulses by a chopper.

[10]The cosine Fourier transform here and the exponential Fourier transform of (38) or (39) are only identical under the conditions used to derive inelastic scattering in section 3. Then $S(Q,\omega)$ is symmetric and $I(Q,t)$ is a real function. The condition $k_{\mathrm{B}}T \gg \hbar\omega$ is fulfilled here because soft matter experiments are done around room temperature and NSE is used to detect small energy transfers.

A variant of the NSE technique described here replaces the main precession coils with a static field by smaller radio-frequency driven coils (neutron *resonance* spin echo, NRSE) [50]. It is even possible to modify this set-up in a way that it does not need the encoding of information into the neutron spin anymore but uses a modulation of intensity (modulation of intensity by zero effort, MIEZE) [51]. In this way the mentioned problems for spin-incoherent scattering can be avoided.

Appendices

A Elementary properties of the Fourier transform

In this section only those properties of the Fourier transform will be recapitulated which are relevant for the understanding of the results on scattering presented in the lecture[11]. The proofs of the theorems will mostly only be sketched. For the mathematically more inclined reader refs. 53–55 are recommended.
According to the dominant standard in neutron scattering the Fourier transform will be defined as

$$F(X) = \int_{-\infty}^{\infty} f(x) \exp(\mathrm{i}Xx)\mathrm{d}x\,. \tag{97}$$

Wherever possible I will use the functional notation

$$\mathcal{F}_{X|x}[f(x)] = \int_{-\infty}^{\infty} f(x) \exp(\mathrm{i}Xx)\mathrm{d}x\,. \tag{98}$$

to abbreviate the Fourier integral. The inversion of the FT (97) is:

$$f(x) = \mathcal{F}^{-1}_{x|X}[F(X)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(X) \exp(-\mathrm{i}Xx)\mathrm{d}X\,. \tag{99}$$

The proof that (99) is the inverse of (97) is not simple and can be found in ref. 53.
A straightforward generalisation of the FT is that to multiple dimensions. The most important case of three dimensions is:

$$F(X,Y,Z) = \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\int_{-\infty}^{\infty}\mathrm{d}z \exp(\mathrm{i}Xx)\exp(\mathrm{i}Yy)\exp(\mathrm{i}Zz) f(x,y,z)\,. \tag{100}$$

The arguments of the exponentials can be grouped together and replaced by the scalar product of the vector $\mathbf{r} = (x,y,z)$ and the 'reciprocal space' vector $\mathbf{Q} = (X,Y,Z)$:

$$F(\mathbf{Q}) = \int \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r}) f(\mathbf{r})\mathrm{d}^3r\,. \tag{101}$$

The inversion is obviously

$$f(\mathbf{r}) = \frac{1}{(2\pi)^3}\int \exp(-\mathrm{i}\mathbf{Q}\cdot\mathbf{r}) F(\mathbf{Q})\mathrm{d}^3r \tag{102}$$

[11] For a more in-depth introduction into the properties of Fourier transforms relevant for scattering see lecture I of the JCNS Neutron Scattering Laboratory Course 2008 [52].

with one $1/2\pi$ pre-factor for each of the individual coordinates' transform.
An important special case of the 3D FT is that of isotropic functions, where the functions only depend on the absolute values $r = \|\mathbf{r}\| = \sqrt{x^2+y^2+z^2}$ and $Q = \|\mathbf{Q}\|$. In spherical polar coordinates (101) reads:

$$F(Q) = \int_0^\infty \mathrm{d}r \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \; r^2 \sin\theta \exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r}) f(\mathbf{r}) =$$

Because of the symmetry one can assume $\mathbf{Q} = (0,0,Q)$ without loss of generality. Then $\mathbf{Q}\cdot\mathbf{r} = Qr\cos\theta$. In addition, because $f(\mathbf{r})$ does not depend on ϕ the integral over this angle can be carried out: $\int_0^{2\pi}\mathrm{d}\phi = 2\pi$. With this we get:

$$= \int_0^\infty \mathrm{d}r \int_0^\pi \mathrm{d}\theta \; 2\pi r^2 \sin\theta \exp(\mathrm{i}Qr\cos\theta) f(r) =$$

Here, we substitute $\cos\theta \to t$ with the consequence that $\mathrm{d}t = -\sin\theta\mathrm{d}\theta$:

$$= -\int_0^\infty \mathrm{d}r \; 2\pi r^2 \int_1^{-1} \mathrm{d}t \exp(\mathrm{i}Qrt) f(r) =$$

Because only $\exp(\mathrm{i}Qrt)$ depends on t the integration can be carried out resulting in

$$= \int_0^\infty \frac{\sin(Qr)}{Qr} f(r) 4\pi r^2 \mathrm{d}r = \frac{4\pi}{Q}\int_0^\infty f(r) r \sin(Qr)\mathrm{d}r\,. \tag{103}$$

Similarly it can be shown that the inverse FT is

$$f(r) = \frac{1}{2\pi^2 r}\int_0^\infty F(Q) Q \sin(Qr)\mathrm{d}Q\,. \tag{104}$$

Often the function to be transformed, $f(x)$, corresponds to a physical observable and thus is a real function $f(x) \in \mathbb{R}$. It can be easily shown from the definition (97) that in this case

$$\mathcal{F}_{X|x}[f(-x)] = \mathcal{F}_{-X|x}[f(x)] = \left(\mathcal{F}_{X|x}[f(x)]\right)^* \tag{105}$$

where the star denotes the complex conjugate defined as $(a+b\mathrm{i})^* = a - b\mathrm{i}$. In words: The Fourier transform of the mirror image of a *real* function is the complex conjugate of the original function. This implies that the Fourier transform of an even function is real and that of a real function is even.
In order to describe point scatterers (e.g. nuclei being orders of magnitude smaller than the scattered wave) it is convenient to introduce the delta 'function' as the following limit:

$$\delta(x) = \lim_{a\to 0}\begin{cases} 1/a & \text{for } |x| < a/2 \\ 0 & \text{everywhere else} \end{cases}\,. \tag{106}$$

When the limit is carried out the delta function is everywhere zero except for $x = 0$ where it has an infinite value. What is important is that the area under the function is always one during the whole limiting process. Therefore, it can be concluded that $\int_{-\infty}^{\infty}\delta(x)\mathrm{d}x = 1$. This integral condition and the fact that only the value at $x = 0$ does not vanish are the only characteristics of the delta function. Therefore, different definitions by limits are possible, e.g. by narrowing

Gaussian functions. It is somehow justified to say that the delta function is the Black Hole of mathematics, where every individuality of the function is lost as that of a star when it collapses. The property making the delta function interesting is that it is able to pull out a single value of a function from a definite integral:

$$\int_{-\infty}^{\infty} \delta(x) f(x) \mathrm{d}x = \lim_{a\to 0} \frac{1}{a} \int_{-a/2}^{a/2} f(x) \mathrm{d}x =$$

Because in the limit of small a the function does not vary much over the interval $[-a/2, a/2]$ it can be replaced by its value at the centre, $f(0)$:

$$= \lim_{a\to 0} \frac{1}{a} a f(0) = f(0)\,. \tag{107}$$

From the basic property of the delta function (107) follows that the FT as the integral (97) is

$$\mathcal{F}_{X|\,x}[\delta(x)] = \int_{-\infty}^{\infty} \exp(\mathrm{i}Xx)\delta(x)\mathrm{d}x = 1\,, \tag{108}$$

i.e. the FT of the delta function is the constant function $F(X) = 1$. By writing down the inversion formula (99) one obtains that the iFT of a constant is a delta function:

$$\mathcal{F}^{-1}_{x|\,X}[1] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}Xx)\mathrm{d}x = \delta(x)\,. \tag{109}$$

(At this point it is important not to forget the 2π !) Because in the last two formulae $\mathrm{i}Xx$ can be replaced by $-\mathrm{i}Xx$ it is of course also true that the iFT of the delta function is a constant $(1/2\pi)$ and the FT of a constant a delta function multiplied by 2π:

$$\mathcal{F}^{-1}_{x|\,X}[\delta(X)] = \frac{1}{2\pi}\,, \tag{110}$$

$$\mathcal{F}_{X|\,x}[1] = 2\pi\delta(x)\,. \tag{111}$$

There are a couple of theorems on definite integrals including FTs which in physics are usually called 'sum rules'. The simplest of these is that for the infinite integral of the FT itself:

$$\begin{aligned} \int_{-\infty}^{\infty} \mathcal{F}_{X|\,x}[f(x)]\mathrm{d}X &= \int_{-\infty}^{\infty} \mathrm{d}X \int_{-\infty}^{\infty} \mathrm{d}x \exp(iXx) f(x) \\ &= \int_{-\infty}^{\infty} f(x) 2\pi\delta(x)\mathrm{d}x = 2\pi f(0)\,. \end{aligned} \tag{112}$$

(using that $\int_{-\infty}^{\infty} \mathrm{d}x \exp(iXx)$ is the FT of constant 1, see (111).) Here, it is assumed that the integrals over X and x are interchangeable which may cause problems for certain functions. Often it may be necessary to interpret the integral over X as Cauchy principal value and handle discontinuities in $f(x)$ by imposing Dirichlet's condition, $f(a) = (\lim_{x\searrow a} f(x) + \lim_{x\nearrow a} f(x))/2$. The complementary sum rule is even simpler to derive:

$$\mathcal{F}_{0|\,x}[f(x)] = \int_{-\infty}^{\infty} f(x) \exp(\mathrm{i}0x)\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\mathrm{d}x\,. \tag{113}$$

We see that the infinite integral of a FT corresponds to the value of the original function at zero (up to a factor 2π) and vice versa.

The *convolution* of two functions is defined as follows:

$$f \otimes g(x) = \int_{-\infty}^{\infty} f(t)g(x-t)\mathrm{d}t\,. \tag{114}$$

One of the most important theorems on FTs is that the FT of a convolution is the product of the individual FTs:

$$\mathcal{F}_{X|x}[f \otimes g(x)] = \mathcal{F}_{X|x}[f(x)] \cdot \mathcal{F}_{X|x}[g(x)] = F(X)G(X) \tag{115}$$

or vice versa:

$$\mathcal{F}_{X|x}[f(x)g(x)] = \mathcal{F}_{X|x}[f(x)] \otimes \mathcal{F}_{X|x}[g(x)] = F(X) \otimes G(X)\,. \tag{116}$$

Proof of relation (115): From the definition (97) follows straightforwardly:

$$\begin{aligned}
\mathcal{F}_{X|x}[f \otimes g(x)] &= \int_{-\infty}^{\infty} \exp(\mathrm{i}Xx) \left(\int_{-\infty}^{\infty} f(t)g(x-t)\mathrm{d}t \right) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}t \exp(\mathrm{i}Xx) f(t)g(x-t) \\
&= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}t \exp(\mathrm{i}X(x-t)) \exp(\mathrm{i}Xt) f(t)g(x-t) \\
&= \int_{-\infty}^{\infty} \mathrm{d}t \int_{-\infty}^{\infty} \mathrm{d}x \exp(\mathrm{i}X(x-t)) \exp(\mathrm{i}Xt) f(t)g(x-t)
\end{aligned} \tag{117}$$

Substituting $x - t \to x'$ leaves the infinite bounds of the first integral unchanged:

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \mathrm{d}t \int_{-\infty}^{\infty} \mathrm{d}x' \exp(\mathrm{i}Xx') \exp(\mathrm{i}Xt) f(t)g(x') \\
&= \int_{-\infty}^{\infty} \exp(\mathrm{i}Xt) f(t)\mathrm{d}t \int_{-\infty}^{\infty} \exp(\mathrm{i}Xx')g(x')\mathrm{d}x' \\
&= \mathcal{F}_{X|x}[f(x)] \cdot \mathcal{F}_{X|x}[g(x)]
\end{aligned} \tag{118}$$

Note that the critical step of this proof is the exchange of integrations at line (117). For some 'crazy' functions this may not be allowed and in consequence the convolution theorem does not hold. Nevertheless, for 'physical' functions this is usually no issue.
Closely related to the convolution is the correlator (or correlation function) of two functions:

$$\langle f(0)g(x) \rangle = \int_{-\infty}^{\infty} f(t)g(x+t)\mathrm{d}t\,. \tag{119}$$

By a substitution $t \to -t'$ it is easy to show that

$$\langle f(0)g(x) \rangle = f(-x) \otimes g(x) \tag{120}$$

so that the correlator is just the convolution with one of the functions mirrored. Using the convolution theorem and (105) in succession, it follows that the FT of the correlator of real functions is the product of the FTs of the correlated functions with one FT conjugated, i.e.

$$\mathcal{F}_{X|x}[\langle f(0)g(x) \rangle] = \mathcal{F}_{-X|x}[f(x)]\mathcal{F}_{X|x}[g(x)] = F^*(X)G(X)\,. \tag{121}$$

In physics the most important special case is the autocorrelation function where $f = g$:

$$\mathcal{F}_{X|x}[\langle f(0)f(x)\rangle] = F^*(X)F(X) = |F(X)|^2\,. \tag{122}$$

This relation, which expresses that the autocorrelation function $\langle f(0)f(t)\rangle$ and the power spectrum $|F(\omega)|^2$ of a signal $f(t)$ are related by a FT, is known as the Wiener-Khintchine theorem. Only two concrete examples of FTs will be presented here, the exponential/Lorentzian FT pair and the Gaussian function. Interestingly, together with clever use of the rules on Fourier transforms these may cover 90 % of all physical problems. If there is really the necessity to obtain the FT of other functions they may be found in table books as ref. 56. In some cases it may even be more effective to look up the Fourier integral (97) in a table of definite integrals. The (according to the experience of the author) most extensive compilation of such integrals can be found in ref. 57.

Exponential decay: Because the exponential $\exp(-ax)$ diverges for x going to negative infinity one cannot FT it with the two-sided transform. One has either to use a one sided transform or define a function

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \exp(-ax) & \text{for } x > 0 \end{cases} \tag{123}$$

which is cut-off at $x = 0$. This makes also sense in most of the physical contexts, e.g. thinking of the current of a capacitor which is discharged by closing a circuit with a resistor at time zero. The Fourier integral (97) becomes then

$$\int_0^\infty \exp(\mathrm{i}Xx)\exp(-ax)\mathrm{d}x = \int_0^\infty \exp((\mathrm{i}X - a)x)\mathrm{d}x =$$

The indefinite integral of the exponential is known, $\int \exp(Ax) = \exp(Ax)/A$, thus one continues the calculation

$$= \frac{\exp((\mathrm{i}X - a)x)}{\mathrm{i}X - a}\bigg|_0^\infty = \frac{1}{a - \mathrm{i}X} = \frac{a}{a^2 + X^2} + \frac{X}{a^2 + X^2}\,\mathrm{i}\,.$$

The real part (which is usually the physically relevant) is a so-called Lorentzian function which peaks at $X = 0$ and has a full width at half maximum of $2a$. Reverting to the interpretation as a decay in time, a is related to the time constant by $a = 1/\tau$. This means that the decay time and the width of the Lorentzian in frequency are inversely proportional.

Gaussian: The bell-shaped curve of the Gaussian (or 'normal') distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{124}$$

decays to both sides rapidly enough and can thus be FTed two-sidedly. The Fourier integral is:

$$\int_{-\infty}^\infty \exp(\mathrm{i}Xx)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right)\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^\infty \exp\left(\mathrm{i}Xx - \frac{x^2}{2\sigma^2}\right)\mathrm{d}x =$$

In order to simplify the integral we are going to complete the square inside the exponential by adding $\sigma^2X^2/2$ and compensating this by a factor in front of the integral:

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\sigma^2X^2}{2}\right)\int_{-\infty}^\infty \exp\left(-\frac{x^2}{2\sigma^2} + \mathrm{i}Xx + \frac{\sigma^2X^2}{2}\right)\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\sigma^2X^2}{2}\right)\int_{-\infty}^\infty \exp\left(-\frac{(x - \mathrm{i}\sigma^2X)^2}{2\sigma^2}\right)\mathrm{d}x = \end{aligned}$$

The definite integral of the Gaussian distribution always fulfills the normalisation property $(1/\sqrt{2\pi}\sigma)\int_{-\infty}^{\infty}\exp(-(x-A)^2/2\sigma^2)\mathrm{d}x = 1$, irrespectively of its centre A. Therefore, the simple result is:

$$= \exp\left(-\frac{\sigma^2 X^2}{2}\right) . \tag{125}$$

This means that the FT of a Gaussian is a Gaussian, a property which is called 'self-reciprocity'. It is shared by all Hermite functions of which the Gaussian is the simplest [54]. Nevertheless, the Fourier transform is not the *same* Gaussian but one with the reciprocal standard deviation $1/\sigma$ of the original Gaussian. Also the FT is not normalised to one but its maximum is fixed to one. This in turn causes that the area under the FT Gaussian is $\int_{-\infty}^{\infty}\exp(-\sigma^2 X^2/2)\mathrm{d}x = \sqrt{2\pi}/\sigma$.

References

[1] P. Lindner and Th. Zemb (eds.), *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter* (Elsevier, Amsterdam, 2002)

[2] R. J. Roe, *Methods of X-ray and neutron scattering in polymer science* (Oxford University Press, Oxford, 2000)

[3] G. E. Bacon, *Neutron Diffraction* (Clarendon Press, Oxford, 1975).

[4] G. L. Squires, *Introduction to the theory of thermal neutron scattering* (Cambridge University Press, Cambridge, 1978).

[5] S. W. Lovesey: "Theory of Neutron Scattering from Condensed Matter" (Clarendon Press, Oxford, 1984).

[6] M. Bée, *Quasielastic neutron scattering* (Adam Hilger, Bristol, 1988)

[7] T. Brückel, G. Heger, D. Richter and R. Zorn (eds.), *Neutron Scattering* (Forschungszentrum Jülich, 2008, `http://hdl.handle.net/2128/37180`)

[8] W. Brown, *Light Scattering: Principles and Development* (Oxford University Press, Oxford, 1996)

[9] B. Chu, *Laser Light Scattering* (Academic Press, New York, 1974)

[10] H. Z. Cummins and E. R. Pike (eds.), *Photon Correlation Spectroscopy and Velocimetry* (Plenum Press, New York, 1977)

[11] B. J. Berne and R. Pecora, *Dynamic Light Scattering* (Wiley, New York, 1976)

[12] R. Pecora, *Dynamic Light Scattering* (Plenum Press, New York, 1985)

[13] R. Hosemann, A. N. Bagchi: "Direct analysis of diffraction by matter" (North-Holland, Amsterdam, 1962).

[14] L. V. Azaroff, R. Kaplow, N. Kato, R. J. Weiss, A. J. C. Wilson, R. A. Young: "X-ray diffraction" (McGraw-Hill,1974).

[15] O. Kratky, O. Glatter (eds.), *Small Angle X-Ray Scattering* (Academic Press, London, 1982)

[16] R. Klein: "Interacting Colloidal Suspensions" in P. Lindner and Th. Zemb (eds.), *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter* (Elsevier, Amsterdam, 2002), pp. 351–380.

[17] W. van Megen, T. C. Mortensen, S. R. Williams and J. Müller, *Phys. Rev. E* **58**, 6073 (1998).

[18] R. Zorn and D. Richter: "Correlation Functions Measured by Scattering Experiments" in ref. 7, chapter 4

[19] G. Placzek, *Phys. Rev.*, **86** 377 (1952)

[20] J. L. Yarnell, M. J. Katz, R. G. Wenzel, S. H.

[21] G. H. Vineyard, *Phys. Rev.* **110**, 999 (1958)

[22] R. Aamodt, K. M. Case, M. Rosenbaum, P. F. Zweifel, *Phys. Rev.* **126**, 1165 (1962)

[23] P. Schofield, *Phys. Rev. Lett.* **4**, 239 (1960).

[24] A. Rahman, K. S. Singwi and A. Sjölander, *Phys. Rev.* **126**, 986 (1962)

[25] R. Zorn, *Nucl. Instr. Meth. A* **603**, 439 (2009)

[26] `http://en.wikipedia.org/wiki/Static_light_scattering`

[27] B. H. Zimm, *J. Chem. Phys.* **16**, 1093 (1948).

[28] K. S. Schmitz, *An Introduction to Dynamic Light Scattering by Macromolecules* (Academic Press, New York, 1990)

[29] S. W. Provencher, *Makromol. Chem.* **180**, 201 (1979).

[30] D. E. Koppel, *J. Chem. Phys.* **57**, 4814 (1972).

[31] A. Aouadi, M. J. Lebon, C. Dreyfus, B. Strube, W. Steffen, A. Patkowski and R. M. Pick, *J. Phys.: Condens. Matter* **9**, 3803 (1997).

[32] O. Kratky: "Experimental Techniques, Slit Collimation" in [15]

[33] J. S. Pedersen: "Instrumentation for Small-Angle X-ray and Neutron Scattering and Instrumental Smearing Effects" in ref. 1, pp. 127–144

[34] `http://www.esrf.eu/UsersAndScience/Experiments/DynExtrCond/ID16`

[35] G. Grübel, A. Madsen and A. Robert: "X-Ray Photon Correlation Spectroscopy (XPCS)" in R. Borsali and R. Pecora (eds.), *Soft-Matter Characterization* (Springer, New York, 2008)

[36] `http://www.esrf.eu/UsersAndScience/Experiments/SoftMatter/ID10A`

[37] H. Frielinghaus: "Small-Angle Scattering" in ref. 7, chapter 8

[38] G. D. Wignall, D. G. H. Ballard and J. Schelten, *Eur. Polym. J.* **10**, 861 (1973).

[39] R. Zorn: "Inelastic Neutron Scattering: Dynamics of Polymers" in ref. 1, pp. 221–255

[40] M. Monkenbusch: "Time-of-flight spectrometers including NSE" in ref. 7, chapter 10

[41] J. Wuttke: "Inelastic Crystal Spectrometers" in ref. 7, chapter 9

[42] F. Mezei, *Neutron spin echo* (Springer, Heidelberg, 1980)

[43] F. Mezei, C. Pappas, T. Gutberlet (eds.), *Neutron Spin Echo Spectroscopy* (Springer, Heidelberg, 2003)

[44] D. Richter, M. Monkenbusch, A. Arbe and J. Colmenero, *Neutron Spin Echo in Polymer Systems* (Springer, heidelberg, 2005)

[45] F. Mezei, *Communications on Physics* **1**, 81 (1976).

[46] M. C. Bellisent-Funel, R. Daniel, D. Durand, M. Ferrand, J. L. Finney, S. Pouget, V. Reat and J. C. Smith, *J. Am. Chem Soc.* **120**, 7347 (1998).

[47] M. Ohl, M. Monkenbusch, T. Kozielewski, B. Laatsch, C. Tiemann and D. Richter, *Physica B* **356**, 234 (2005).

[48] `http://www.ill.eu/instruments-support/instruments-groups/instruments/in15/`

[49] `http://neutrons.ornl.gov/instruments/SNS/NSE/`

[50] R. Gähler and R. Golub, *Physica B & C* **49**, 1195 (1988).

[51] M. Koppe, M. Bleuel, R. Gähler, R. Golub, P. Hank, T. Keller, S. Longeville, U. Rauch and J. Wuttke, *Physica B* **266**, 75 (1999).

[52] R. Zorn: "Fourier Transforms" in T. Brückel, G. Heger, D. Richter, R. Zorn (eds.): "Neutron Scattering" (Forschungszentrum Jülich, 2008, `http://hdl.handle.net/2128/37180`), chapter I.

[53] K. B. Howell: "Principles of Fourier Analysis" (Chapman & Hall/CRC, Boca Raton, 2001).

[54] H. J. Weaver: "Theory of discrete and continuous Fourier analysis" (John Wiley & Sons, New York, 1988).

[55] J. Kauppinen, J. Partanen: "Fourier Transforms in Spectroscopy" (Wiley-VCH, Berlin, 2001).

[56] F. Oberhettinger: "Tables of Fourier Transforms and Fourier Transforms of Distributions" (Springer, Berlin, 1990).

[57] A. P. Prudnikov, Yu. A. Brychkov, O. I. Marichev: "Integrals and Series" (Gordon and Breach, New York 1986–1992).

C 2 Light Microscopy

R. Merkel
Institute of Complex Systems, ICS-7
Forschungszentrum Jülich GmbH

Contents

1 Introduction 2

2 Ray optical description of the light microscope 2

3 The resolution of the light microscope 6

4 Diffraction effects in imaging 8

5 Fluorescence microscopy 9

References 12

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

Optical microscopy is one of the most powerful techniques in life sciences. It is also used very often in soft matter sciences. The reasons for the tremendous potential of light microscopy are:

a) The weak interaction between visible light and (most) materials. Therefore biological molecules and structures experience negligible stress during observation. Moreover, soft samples are barely disturbed by optical observation.
b) A good spatial resolution of about 250 nm, well suited to many biological and soft matter research questions.
c) The availability of many contrast methods.
d) The availability of very bright and well controllable light sources (lasers, high power light emitting diodes, classical light bulbs and high pressure arc lamps) together with very sensitive and highly accurate light detectors like solid state cameras (CCD and CMOS) or photomultipliers.

In recent years the progress of computation power has made digital image processing on micrographs captured by digital camers a routine procedure. This combination is often called "quantitative microscopy" and has enabled a plethora of quantitative analyses of microscopic phenomena. Today processes like fluctuations of biopolymers, shape changes of membranes or diffusion of single molecules can be quantitatively analyzed using light microscopy. Last but not least, major parts of the works resulting in last year's Nobel Prize in chemistry for superresolution microscopy were based on quantitativel microscopy

The importance of the light microscopy and its rapid development is demonstrated by a literature search in ISI Web of Science for papers with [(optical or light) and microscopy] in their topics. This resulted in 47012 hits alone for the year 2014.

Moreover, the field of cell biology is effectively defined by light microscopy. Organisms too small to be studied with this technique are classified as topics for microbiology instead of cell biology.

Surprisingly, light microscopy is often neglected in physics, biology and chemistry curricula. This lecture tries to fill this gap focusing on basic principles. I will start with a description of light microscopy in terms of ray optics, discuss shortly optical resolution and will end with some important contrast methods. Luckily there are many books on light microscopy, for example see [1-4].

2 Ray optical description of the light microscope

The best starting point to understand the function of a microscope is a thin lens, cf. Fig. 1. Almost all geometrical considerations on image formation by thin lenses are rested on three principles: a) the beam path can be reverted; b) a parallel beam of light is focused into a point in the focal plane; and c) a beam of light passing through the center of a thin lens is not refracted.

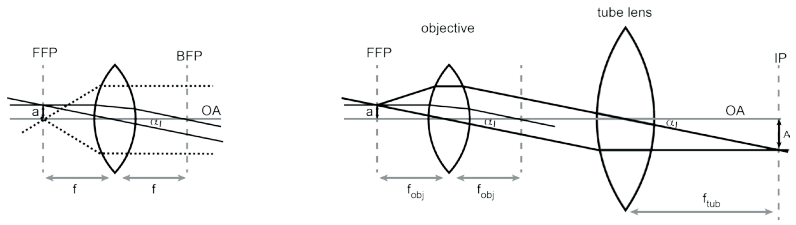


Fig. 1: *Imaging by a thin lens (left) and the image sided beam path of a microscope (right). a object size, A image size, f focal length, OA optical axis, FFP front focal plane, BFP back focal plane, IP image plane. Note that a real objective is composed of many lenses and therefore its front focal length is not identical to its back focal length.*

An object placed into the focal plane of a lens (cf. Fig. 1, left) is converted into parallel bundles of light. In other words, its image is at an infinite distance from the lens. For an object of size a the opening angle α of the cone of light beams produced fulfills $a = f \tan\alpha$. Please note that the angle of incidence of a parallel beam of light with respect to the optical axis of the lens corresponds to the location of the point in the focal plane it is imaged on and vice versa. In other words, in imaging to infinity angles of beams and locations of image points are converted into each other.

With this in mind we can approach the image sided light path of the microscope. Modern light microscopes are compound microscopes, i.e. they consist of separate optical elements. Since the mid to late 70s almost all research microscopes are designed according to the so-called "infinity" optics. Here, the object is placed in the front focal plane of the microscope objective. Thus it is imaged at infinity. An real image is again created by the tube lens (cf. Fig. 1, right).

Because the central ray through a thin lens is not refracted, object and its image are both imaged by rays of the same angle α (cf. Fig. 1, right). Thus basic geometry tells us the magnification M of the infinity optics microscope:

$$M = \frac{A}{a} = \frac{f_{tub}}{f_{obj}} \tag{1}$$

Typical values for the magnification range from 2.5 to 100. Different manufacturers use different focal lengths of the tube lens. A typical value is 200 mm. Using an infinity corrected objective of one manufacturer on a microscope body of another is often well possible albeit often with altered magnification.

The space between objective and tube lens is often called "infinity space". Here light rays originating from one object point are parallel and the distance between objective and tube lens is almost arbitrary. Therefore, the lens can be moved for focussing while the rest (microscope body and sample) remains stationary. Especially for complex sample environments (e.g. electrophysiological measurement under light microscopic control) this is a major advantage compared to the traditional system where the sample stage is moved for focussing. Moreover, optical elements like filter plates and polarizers can be introduced into the infinity space

without altering focus position. This again is a major advantage for observing the same sample with different optical techniques (see below).

To summarize, the optical principle of the infinity optics is as follows. The object is in the front focal plane of the objective, the "image" at infinity. A real image is created by the tube lens at its focal plane.

Before the advent of modern cameras the real image created by the tube lens was again projected towards infinity by the eyepiece. This enables observation with an infinity adapted eye. Nowadays the camera is either placed in the focus plane of the tube lens or (most often) an additional optics is used to project the image onto the camera.

An often underrated optical part of the microscope is the condenser. However, as demonstrated in Fig. 2, it is of almost equal importance for clear and well contrasted images as is the image sided beam path. The design of the condenser must solve three problems. First, the objective is only capable to observe a certain area on the object, the so-called field of view. Illumination of other areas can only result in additional stray light and unnecessary sample degradation. Second, the objective can collect light only within a cone (cf. Fig. 2, left). Light at inclinations outside this cone will hit lens fittings or other construction elements of the objective and, again, result in stray light. Third, the field of view should be illuminated with constant intensity.

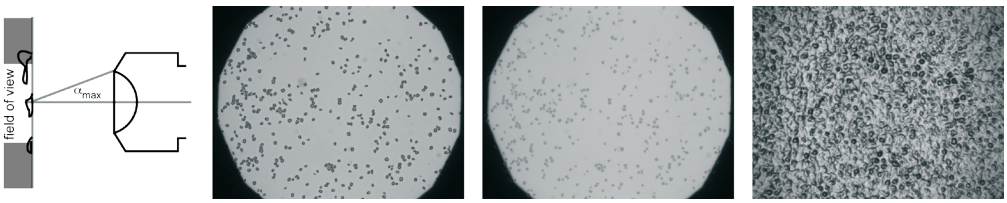


Fig. 2: *The importance of the illumination for image quality. From left to right: (i) Sketch of the requirements for immumination. (ii) Correct image (sample: Clamydomonas rheinhardtii, objective with 5x magnification and 0.15 numerical aperture, angles of illumination and observation are matched). (iii) Illumination with 3.5x larger angles than accepted by the objective results in severe loss of contrast. (iv) An uneven illumination (produced by defocussing the condenser until a ground glass screen became visible) severely reduces clarity of images.*

Taken together an ideal condenser should illuminate a defined area with a cone of light of defined opening angle. This optical challenge is solved by the Köhler optics, cf. Fig. 3.

It is again based on infinity imaging. A real image of the light source is created by the collector lens. In addition to the basic elements shown in Fig. 3, most often a heat absorbing filter and a ground glass screen to homogenize the light are used between light source and collector. This image of the light source is placed in the front focal plane of the condenser lens. Thus it is imaged to infinity. The aperture stop is also placed in the image plane of the light source. It is used to regulate the size of the light source. The central point of the light source creates a parallel beam of light in direction of the optical axis, off-axis points create inclined beams of parallel light. The maximum inclination α_{max} of light beams leaving the

condensor is given by the radius r of the condenser lens and the focal length of the condenser lens via $\tan\alpha_{\max} = r / f_{cond}$.

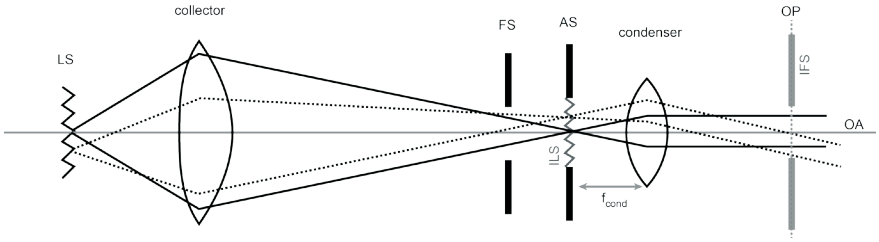


Fig. 3: *Condenser according to Köhler. LS: light source (filament), FS field stop, AS aperture stop, OP oject plane, OA optical axis, ILS image of light source, IFS image of field stop. Light emitting from an axial (full lines) and an off-axis point (dotted lines) is depicted.*

While the placement of the aperture stop at the focal plane of the condenser lens solves the angular constraints, the field of view is defined by placing a field stop at a location that is imaged onto the object plane by the condeser lens.

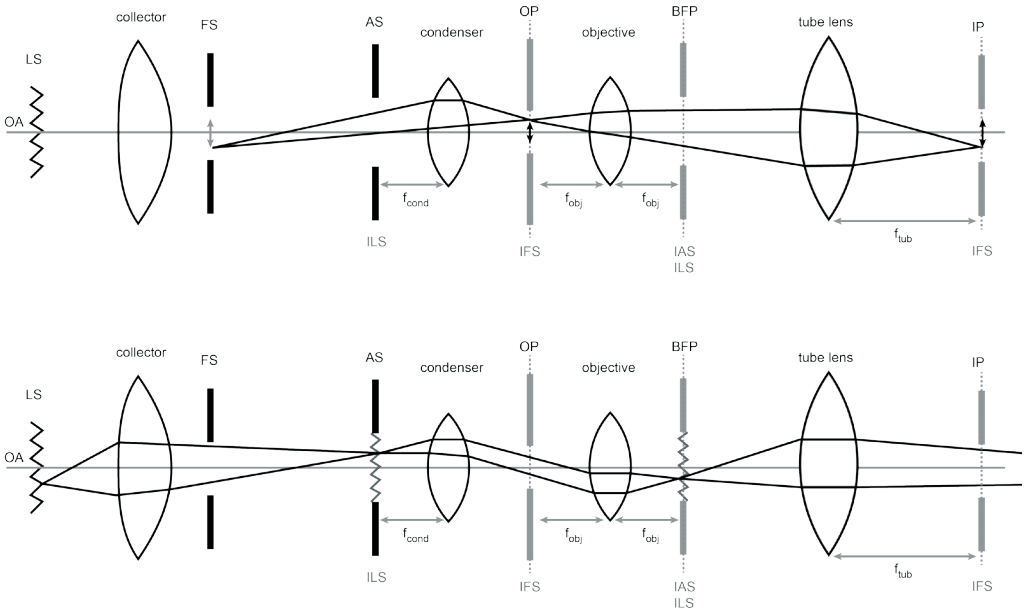


Fig. 4: *Object beam path (top) and illmination beam path (below) of an infinity optics microscope. OA optical axis, LS light source, FS field stop, AS aperture stop, OP object plane, BFP back focal plane of objective, IP image plane, ILS image of light source, IFS image of field stop, IAS image of aperture stop.*

Homogeneity of illumination is reached by imaging the light source to infinity, i.e. it is maximally defocused in the object plane. Any other location of the light source would

produce a defocused image of it within the object plane. Therefore proper alignment of the condenser is mandatory for good quality micrographs.

The optical design of the light microscope can be depicted in two ways. Either one follows the images of the object ("object beam path") or those of the light source ("illumination beam path"). Wherever the object is focussed the light source must be maximally defocussed (i.e. imaged to infinity) and vice versa. While this sounds simple, most problems in practical use or technical modification of a light microscope are caused by failure to follow this simple principle. Both beam paths are displayed in Fig. 4 which summarizes the ray optical design of the infinity optics light microscope.



Fig. 5: *Practical implementation of infinity optics. Left: An upright microscope; Right: an inverted microscope.*

Research microscopes are of modular design enabling different ways of illumination (throught the object or in reflection), change of objectives, the use of additional optics in the infinity space and observation with the eye or different cameras, cf. Fig. 5. The optical elements shown in Figs. 1-4 as simple thin lenses are in reality combinations of several ones to compensate for imaging defects (lens aberrations). Moreover, the mechanical precision of all parts of the microscope body must be extremely high to enable reliable performance. As many scientists have found the hard way, building a working microscope is quite a demanding task. If you work with a microscope: Respect the skill that went into its construction and manufacturing and handle it with care. Misuse and dirt are the foremost enemies of the microscope (and your future measurements).

3 The resolution of the light microscope

Ray optics is sufficient to understand the basic function of a compound microscope but cannot derive the resolution limit of the microscope. Therefore we must take into account the wave nature of light. For a first approach imagine the assembly of objective and tube lense, cf. Fig. 6. The objective collects light within a cone with opening angle amax from an object point. This point emits a spherical wave train. In the end the whole function of the objective and tube lens is altering the wave front in a way that it again forms a spherical wave front focusing in the image point.

Modern objectives and lenses are so well corrected that aberrations (i.e. deviations from this behavior) play little role. The optical image of a light point in the object is calculated by summing up the interference of converging rays emerging from a circular aperture the size of which is given by α_{max} and the focal length of the lens. The final result is given by [2,5].

$$I_{point} = \left[2\,\mathrm{J}_1(\xi)/\xi\right]^2 \text{ with } \xi = \frac{2\pi}{\lambda} n \sin\alpha_{\max} \frac{r}{M} \tag{2}$$

Where J_1 denotes the first Bessel function of first kind, l the wavelength of light, n the refractive index in the object space, r the radial distance in the object plane and M the magnification. The dimensionless variable ξ is often described as "image position in optical units". On the basis of these equations the numerical aperture, NA, of an objective is defined as

$$NA = n \sin\alpha_{\max} \tag{3}$$

The angular spread of light emerging from the condeser is described by the "illumination numerical aperture" INA that is defined analogously.
Within the object plane the "optical unit" is given by $\lambda/2\pi NA$. The minima of the image of a point occur at x=3.83, 7.02, 10.17, ... and the maxima at 0, 5.14, 8.46, 11.62, ...

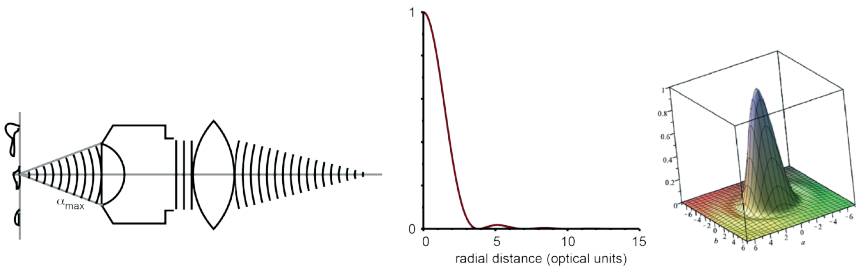


Fig. 6: *Left: Scheme of light waves during imaging. Middle: Radial intensity distribution of the image. Right: 3D representation of the intensity distribution of a point.*

The image of a point source is also called "point spread function (PSF)". This quantity is of paramount important in the interpretation of light microscopic images. As the image may be thought of as summation of all point images the final image of an object is a convolution of the "true" image with the PSF. Here, we have shown only the PSF in the image plane. This quantity is often also called the "Airy disc". However, the PSF is a three-dimensional function that also describes the blurred image of an out-of-focus object.

Exact calculation of the PSF in bright field microscopy is a demanding task as it is influenced by the spatial coherence of the light source and the illumination numerical aperture (Pluta). Here we will not enter this complicated aspect.

Finally the PSF can be used to define the resolution limit of a light microscope. If two incoheren, self-luminous objects are approaching them to the point where the maximum of one Airy disc is just at the location of the first minimum of the second, both can still be separated. If the distance is closer, separation is no longer possible. This consideration results in a resolution limit of [2,5]

$$\Delta r_{\min} = \frac{0.61\lambda}{NA} \tag{4}$$

This equation shows that wavelength and numerical aperture of the objective limit microscopic resolution. Imaging is done usually with visible light (400 nm to 750 nm) where blue light is often already harmfull to living objects. The numerical aperture of an air lens is limited to 0.95 as the opening angle of the observation light cone cannot reach the full 90°. In water immersion numerical apertures of 1.25 are possible, with oil immersion this value raises to 1.45. Therefore resolutions of down to 170 nm are possible in classical widefield microscopy.

4 Diffraction effects in imaging

An alternative approach to optical resolution is considering imaging of gratings. Such a grating of periodicity d will create a classical diffraction pattern with maxima in directions Θ_l satisfying $\sin\Theta_l = l\lambda / d$. A beam of parallel light will emerge from the object in each of these directions. All diffraction orders within the acceptance angle of the objective will be collected and imaged into points in the back focal plane of the objective. These spots are again sources of spherical wave fronts that will be focused by the tube lens into an image of the grating. The resolution of the image obviously improves with the number of diffraction orders collected.



Fig. 7: *Imaging of a grating (10 µm periodicity) with different conditions. All images scaled to the same print magnification, contrast adapted. Left: Objective with 2.5 magnification and NA 0.075 (1 diffration order collected). Middle and right: Objective with 63x magnification and NA 0.75 (15 diffraction orders collected). Middle: Illumination aperture (INA)≈0.3. Right: INA≈0.07.*

As demonstrated in Fig. 7, the resolution of a grating image increases dramatically with increasing number of diffraction orders collected. In addition, the illumination numerical aperture (INA) influences image contrast. The "satellite" fringes seen in Fig. 7 at very low INA are the product of image reconstruction from a finite number of diffraction orders. As is well known from the theory of Fourier series, a finite Fourier sum will reconstruct a square wave with pronounced over- and undershooting around the edges. This gives raise to such diffraction artefacts. A very simple way of dealing with them is opening the condenser aperture.

As clearly seen in Fig. 8, the diffraction pattern observed in the back focal plane is strongly influenced by the condenser. Each luminous point with the aperture stop (cf. Fig. 3) gives rais

to a full diffraction pattern. Patterns originating from off-axis points are shifted according to the angle at which this point is "seen" in the object plane (cf. Fig. 3).

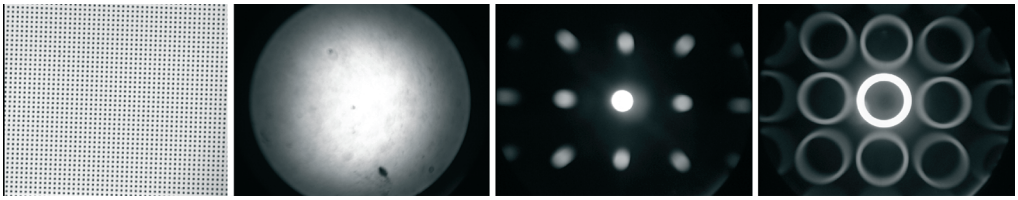


Fig. 8: *The effect of the illumination aperture in the diffraction image in the back focal plane. Left: The imaged square lattice (1.75 µm periodicity), lens 50x magnification, NA 0.7, INA ≈0.3. All others: Intensity in the back focal plane of the objective (diffraction images). First with INA≈0.3, second with INA≈0.07, third illumination with a cone of light.*

Therefore the final diffraction pattern observed in the back focal plane is a convolution of the true diffraction pattern and the angular intensity distribution impinging on the sample. This is most dramatically demonstrated using a cone of light where all diffraction spots are replaced by rings. Clearly separated diffraction orders are seen at very small illumination aperture, whereas at large INA these points will overlap creating a smoothly varying intensity distribution in the back focal plane.

Mathematically the intensity distribution in the back focal plane is the Fourier transform of the image intensity distribution [6,7]. A Fourier transform converts a smoothly varying function in another smooth function whereas functions exhibiting sharp cut-offs have transforms with pronounced oscillations. This effect explains why opening the condensor is reducing diffraction artefacts in the final images. Moreover, the fact that the diffraction image and the real image form a Fourier transform pair can be exploited for filtering of images, impressive examples are shown in [7].

The fact that at low INA diffracted and directly transmitted light (zeroth and higher orders in Fig. 8) are spatially separated in the back focal plane of the objective can be exploited to alter phase or intensity of the transmitted light with respect to the diffracted light. Image contrast techniques like phase contrast or modulation contrast rely on this [2,6].

5 Fluorescence microscopy

Unfortunately, bright field microscopy and all sophisticated techniques based on it are not able to image the distribution of a specific molecule, let's say the insulin receptor or the lipid sphingolmyelin, in cells. This formidable challenge has been partially solved in fixed cells by immunofluorescene microscopy [8] and in living cells by transfection with genetic vectors encoding fusion proteins that comprise the molecule of interest and a fluorescent molecule [9, 10].

In fluorescence microscopy fluorophores, that is, molecular moieties exhibiting large delocalized electron systems capable of light absorption and re-emission are attached to the molecules under investigation. The physical processes occuring in fluorophores are usually

summarized in a Jablonski diagram, cf. Fig. 9. Light of high energy ("blue light") is absorbed by the fluorophore. This causes electronic and vibrational exciation. Vibrational excitations relaxe extremeley fast and together with internal conversion (reduction of the electronic excitation whith simultaneous vibrational excitation conserving energy) quickly bring the molecule into the first electronic excited state (S_1) without vibrations. In most dyes internal conversion is so efficient that this state also rapidly decays to the electronic ground state, effectively converting the whole photonic energy into heat. In contrast, in fluorophores this state can also decay via emission of a photon (fluorescence). This whole process typically occurs on a time scale of nanoseconds. Obviously, fluorescence is red shifted compared to excitation. The situation is complicated by the presence of the triplet state where the electron spins are parallel. As spin-flips are quantum mechanically forbidden, photon emission from the triplet state is very unlikely and often the molecule resides for long time periods (ms to s). As this is a high energy state, a molecule in the triplet state is chemically very active and prone to chemical degradation (photobleaching). Please note that molecular fluorescence is a a rich scientific field. Interested readers should consult [11-15].

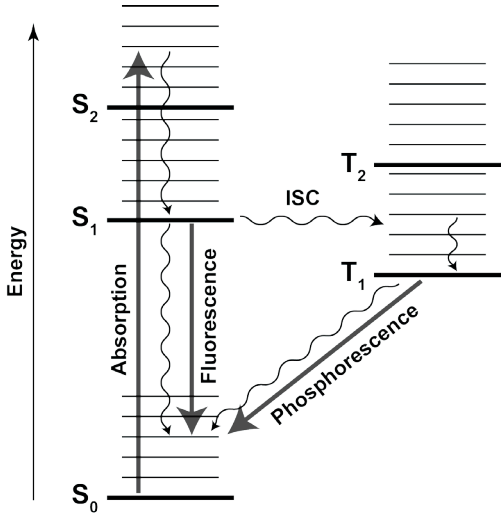


Fig. 9: *Jablonski scheme of fluorescence. Bold arrows: transitions involving photons, wavy arrows: radiationless transitions. Bold lines: electronic states. Thin lines: vibrational states. For further explanation see text.*

As fluorescent light is red shifted it can be easily separated from the excitation light. This is accomplished with a filter cube comprising an excitation filter (transmits only blue light), a dichroic mirror (reflects blue light, transmits red light) and a emission filter (blocks blue light and transmits red light). In most implementations the objective also serves as condenser. However, this simplification comes at the prize of an angled beam path, cf. Fig. 10.

Moreover, because of the finite absorption coefficients of fluorophores and the limited quantum yield of fluorescence, fluorescent light intensities are usually extremely low, orders of magnitude below the excitation intensity. Therefore, very intense light sources are needed. Here, high pressure arc lamps (mercury or xenon), lasers, and, recently also, high power light emitting diodes are used in combination with very efficient optical filters and very sensitive digital cameras or photomultipliers.

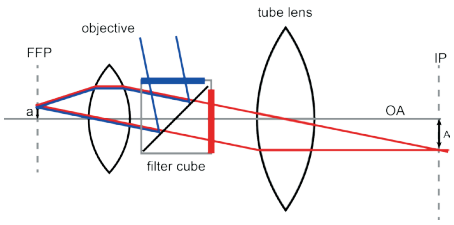


Fig. 10: *Light path of a fluorescence microscope with a filter cube separating excitation (blue) and emission (red).*

Because photon absorption and photon emission are separated by multiple relaxation processes, each fluorophore is a independent, incoherent light source. Therefore, image formation in fluorescence microscopes is much easier to understand than in bright field microscopes where coherence effects are adding diffraction artefacts to the images (cf. Fig.7).

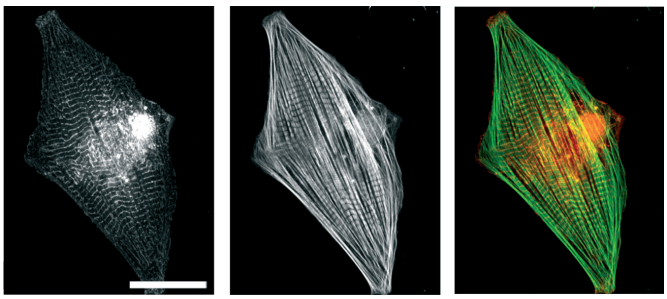


Fig. 11: *Immunfluorescenec microscopy on cultivated heart muscle cells. Left: antibody against α-actinin. Middle: antibody against actin. Right: overlay (α-actinin red, actin green). Scale bar 50 µm. Taken from [16].*

Examples for fluorescence microscographs are shown in Fig. 11. Please note that in the context of cell biology and cell biophysics most microscopy is done in fluorescence.

References

[1] K. Michel, Die Grundzüge der Theorie des Mikroskops in elementarer Darstellung (Wissenschaftliche Verlagesgesellschaft, Stuttgart, 1964)
[2] M. Pluta, Advanced Light Microscopy (Elsevier, Amsterdam, 1988)
[3] G. Sluder, Video Microscopy (Academic Press, San Diego, 1998)
[4] D. G. Murphy, Fundamentals of Light Microscopy and Electronic Imaging (Wiley, New York, 2012)
[5] M. Born and E. Wolf, Principles of Optics (Cambridge University Press, 2009)
[6] E. Hecht, Optics (Addison-Wesley, San Francisco, 2002)
[7] M. Francon, Modern Applications of Physical Optics (Interscience Publ., New York, 1963)
[8] R. H. Coons and M. H. Kaplan, J. Exp. Med. 91, 1 (1950)
[9] M. Chalfie, Y. Tu, G. Euskirchen, W. Ward and D. Prasher, Science 263 (1994)
[10] G. Jung, Fluorescent Proteins (Springer, Berlin, 2012)
[11] J. B. Birks, Photophysics of Aromatic Molecules (Wiley-Interscience, London, 1970)
[12] B. Valeur, Molecular Fluorescence (Wiley-VCH, Weinheim, 2002)
[13] J. N. Turro, J. C. Scaiano and V. Ramamurthy, Modern Molecular Photochemistry of Organic Molecules (University Science Books, Sausalito, 2010)
[14] P. J. Klan and J. Wirz, Photochemistry of Organic Compounds (John Wiley, Chichester, 2009)
[15] J. R. Lakowicz, Principles of Fluorescence Spectroscopy (Springer, New York, 2006)
[16] N. Hersch, B. Wolters, G. Dreissen, R. Springer, N. Kirchgeßner, R. Merkel and B. Hoffmann, Biology Open 2, 351 (2012)

C 3 Single-Molecule Fluorescence Spectroscopy

J. Fitter
Institute of Complex Systems / Molecular Biophysics (ICS-5),
Forschungszentrum Jülich GmbH and
I. Physikalisches Institut (IA), AG Biophysik, RWTH Aachen

Contents

1 Introduction **2**

2 Instrumentation for Optical Single Molecule Detection **3**
2.1 TIRF Microscopy 3
2.2 Confocal Microscopy 4

3 Single-Molecule Data Acquisition and Analysis Methods **6**
3.1 Förster Resonance Energy Transfer (FRET) 7
3.2 Fluorescence Correlation Spectroscopy (FCS) 10
3.3 Fluorescence Lifetime and Anisotropy 12
3.4 Multiparameter Burst Analysis 15

4 Conclusion **16**

5 Acknowledgement **17**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Whenever we interpret our experimental data, for example related to studies on molecular processes, we think and talk about single molecules. However, our experimantal data is typically measured with huge numbers of molecules, called ensembles. How can we be sure that all molecules behave in the same way and that such a concept of assuming that all molecules are described reasonably well by ensemble average observation parameters makes sense at all? The answer is: often we cannot be sure from a priori knowledge. In particular in soft matter or biological systems we deal with pronounced heterogeneity and complex behaviour. Therefore the study of single molecules, which is now possible by the use of some recently developed techniques, offers the most straightforward way of testing this assumption. Single molecule studies provide an insight into the behaviour of each individual molecule and thereby disclose the detail of subpopulations in structure and dynamics within an ensemble. It is important to note, that in single molecule experiments we typically characterize hundreds or thousands of individual molecules, which gives us access to the most valuable determinant of our measurements: the width and the shape of the distribution of our property of interest (see Figure 1). In addition, single molecule methods provide a way of probing the temporal evolution of molecule parameters without the need for synchronization. Furthermore, single molecule techniques can also provide a very straighforward comparision with theory and with the results of computer simulations [1, 2].

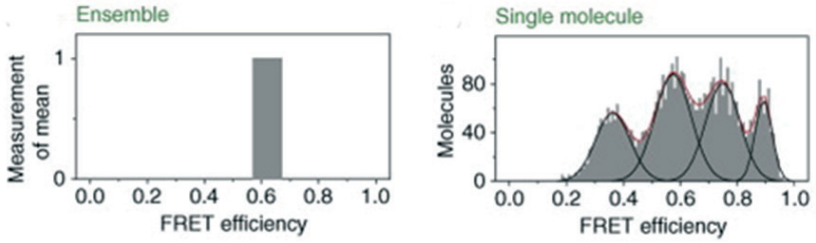


Fig. 1: *Here the difference in the obtained molecule property parameter for an ensemble and a single molecule measurement is shown. While in ensemble measurements we only obtain a mean value (FRET efficiency value of 0.6, for details about this parameter see Subsection 3.1), single molecule data povide details about the distribution of the parameter, including the statistical weight and the distribution width of all subpopulations. From Ref. [3]*

Besides force based methods (AFM, optical and magnetic tweezers) and the recently developed single particle cryo electron microscopy, fluorescence spectroscopy and microscopy is the most important single molecule technique for life science applications. As we will see in Section 3, various methodical approaches and manifold applications with biomacromolecules gave rise to the fact that the fluorescence based single molecule methodology is going to become more and more a common tool in molecular biology [4, 5, 6, 7, 3, 8]. Considering optical studies, fluorescence techniques benefit from the fact that the rather small number of photons which are red-shifted with respect to a huge number of photons in the incident light can be separared very effectively from the later. As a consequence signal-to-noise ratios, appropriate to allow single molecule detection, can be achieved when sample backgrounds are carefully controlled. A detailed description of how optical setups are designed in order to reach single molecule sensitivity is given in the following Section.

2 Instrumentation for Optical Single Molecule Detection

In general we are dealing with two main types of fluorescent measurements that will be discussed in the following: measurements on (freely) diffusing fluorescent single molecules and measurements on immobilized single molecules. The far most common optical setups are based on commercial microscopes employing the inverted epi-fluorescence configuration [9, 2]. Among those the Total Internal Reflection Fluorescence (TIRF) microscope and the confocal microscope will be described here in more detail. In both configurations one major goal is to restrict the excitation light to small volumes which is the key in achieving reasonable signal-to-noise levels.

2.1 TIRF Microscopy

In TIRF microscopy, like in classical wide-field microscopy, we image fluorescent single molecules at the resolution limit. Using this kind of setup (Figure 2), on the one hand only immobilized or slowly moving molecules can be measured, but on the other hand, many (up to hundreds) molecules can be visualized simultaneously over extended observation times (up to tens of seconds) with submillisecond time resolution [5, 7].

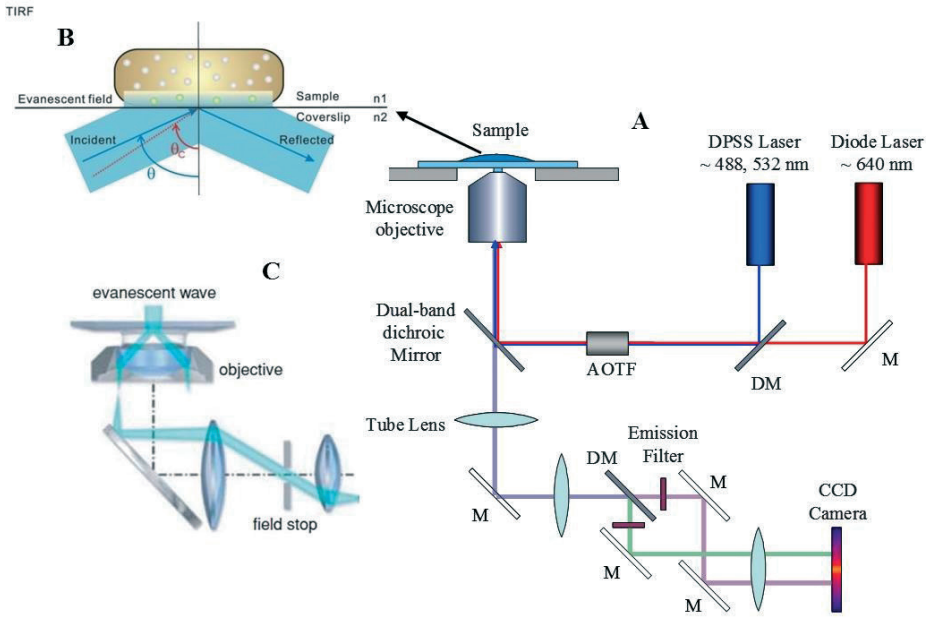


Fig. 2: *(A) Here a simple scheme for simultaneous dual color detection is shown, as it is employed in wide field or TIR illumination (optical elements: DM: dichroic mirror; M: mirror; AOTF: acousto-optical tunable filter). (B) In TIR illumintaion (here the through-the-objective configuration is employed, see C) an evanescent field is generated at the interface between the coverslip and the sample. Only in this volume, with a penetration depth smaller than approximately 150 nm into the sample regime, excitation of fluorescent molceules takes place. (C) In order to achvieve excitation light at incident angles larger than the critical angle, the excitation beam is shifted off-axis. Panel (A) from Ref. [8]*

In wide field or TIRF microscopy light from cw lasers is reflected by the major dichroic mirror, or by a dual-band dichroic mirror if simultaneous two-color excitation is used, into a microscope objective. A convenient approach of creating TIR illumintaion is achieved by focussing a collimated beam onto the periphery (off-axis) of the back focal plane of a high numerical aperture microscope objective (through-the-objective configuration, Fig. 2C). By moving the excitation beam away from the on-axis position to a more or less pronounced off-axis position, the angle of incidence reaches and even exceeds the critical angle of total reflection. This creates an evanescent field, penetrating into the sample volume by a very limited depth (Fig. 2B). To achieve TIR the numerical aperture (NA) of the objective should be as large as possible, accessible for example with water or oil immersion objectives. Useful NA values of ~ 1.45 give a maximum theoretical angle of incidence of 73.8°, which is well above the critical angle of total reflection (θ_c = 63° for a glass-water interface). Another common, but maybe less convenient, method for achieving TIR is to use a prism (for more details see [2]).
The fluorescence emission light is thereafter collected by the same objective and, due to its red-shift with respect to the excitation light, able to pass the major dicroic mirror. Since a certain area of the sample (field of view) is illuminated with the excitation light, the tube lense facilitates the image formation of the diffraction limited fluorescence peaks, originating from molecules close to the cover slide surface, on the image sensor of a CCD camera. For many applications (e.g. co-localisation of different species of molecules or FRET, see Subsection 3.1) a setup with at least two colors, which need to be measured simultaneously, is required. Therefore, in the dual-color setup, as shown here, the image of both colors are split by a further dichroic mirror and finally projected into two different areas of the camera surface (see lower part of Fig. 2A). In typical applications, intensity trajectories extracted from many subsequent images are analysed to monitor slow lateral movements or the time course of molecular processes of surface-immobilized molecules [5, 7].

2.2 Confocal Microscopy

In contrast to TIRF microscopy the confocal microscope can facilitate measurements with surface tethered molecules and with molecules diffusing in solution. This property is achieved by the fact that in confocal microscopy the fluorescence detection volume is limited in all three spatial directions. Due to the fact that the excitation light is tightly focused to a diffraction limited spot in the sample volume and that a confocal pinhole is resticting fluorescence emission to a well defined focal plane perpendicular to the optical axis (Fig. 3C) we obtain effectively a very small detection volume (Fig. 3B). For light in the visible regime and pinholes with an aperture diameter of 20-70 microns this detection volmune is characterized by a radial extension of a few hundred nanometer and an axial dimension of a about 2-4 microns.
As in wide-field and TIRF microscopy the confocal microscopy requires high NA values ($\sim 1.2 - 1.4$) for the employed objectives, not only to achieve the high spatial resolution, but also because of the high photon collection efficiency for fluorescence light determined by the large objective aperture angles. The later is particularly crucial for single molecule sensitivity. Corresponding to the way of sample illumination the fluorescence light is detected by a single-point detector (Fig. 3A). These detectors (e.g. avalanche photo diodes) are characterized by a rather high detection efficiency (50-80 $\%$), but in contrast to CCD cameras with a much higher temporal resolution in the order of a few hundred picoseconds. This property enables fluorescent lifetime measurements and extended use of photon correlation techniques (see Subsection 3.3 and 3.2). At least for the lifetime applications pulsed excitation is necessary. A very powerful

approach to obtain lifetime information (for fluorophores with fluorescent lifetimes in the order of a few nanoseconds) is to employ time correlated single photon counting (TCSPC), see for example [10]. For this purpose the laser sources should exhibit pulse widths shorter than 500 picoseconds and have to be operated at repetition rates ranging from 10 - 80 MHz.

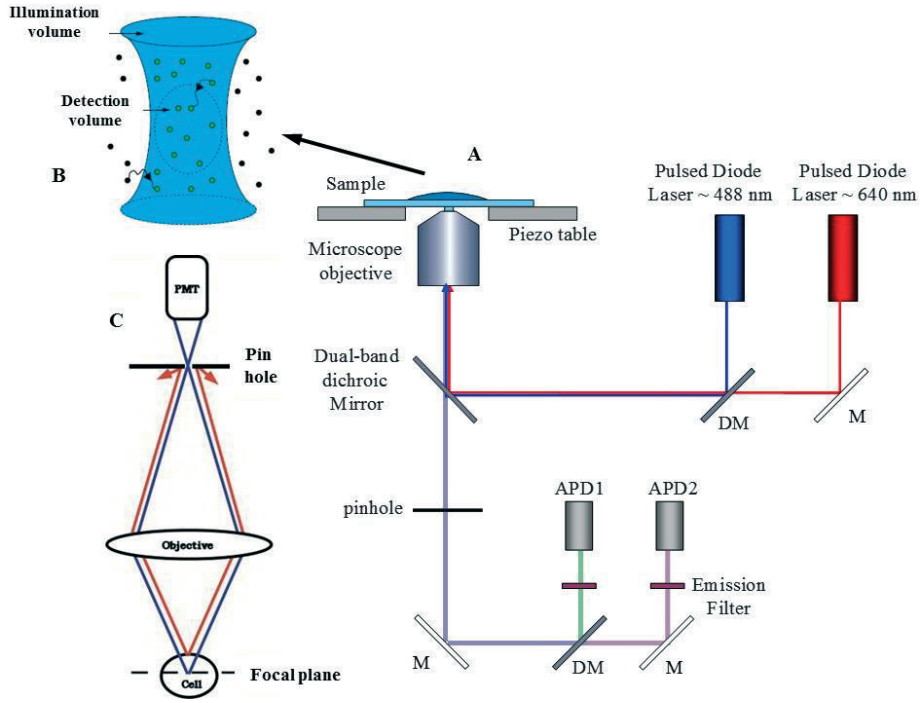


Fig. 3: *(A) In this figure the scheme of a confocal microscope with dual color detection is shown (legend for optical elements see Figure 2). Here the photon detection is accomplished by avalanche photo detectors (APD). Therefore this setup can also be used for measuring fluorescence lifetimes, if pulsed excitation is employed. (B) In the confocal setup the excitation light is focused to a diffraction limited spot which results in very small detection volumes. These small detection volumes can be utilized for image formation by raster scanning approaches or by measuring freely diffusing molecules in highly diluted solutions (shown here). (C) The restriction of the detection volume along the optical axis is achieved by using a confocal pinhole as depicted here. Panel (A) from [8]*

Similar to the TIRF microscope the confocal setup can also detect two colors at the same time. In order to achieve dual color detection the fluorescence light is separated by a further dichroic mirror after the pinhole and finally transmitted through respective emission filters onto the active surface of the APDs.

One major application of confocal microscopy is to detect fluorescent photons from a small number of molecules (or even of individual molecules) in solution passing the detection volume driven by Brownian motions. A very powerful approach to measure precisely the concentration of molecules within the detection volume and to determine the diffusivity of fluorescently labeled molecules is given by fluorescence correlation spectroscopy (FCS, see Subsection 3.2). In order to obtain reliable quantitativ results from FCS, the size and the shape of the detection

volume has to be known a apriori. To achieve this, FCS applications are typically performed by using excitation light beam diameters which are smaller than the back focal aperture diameter of the microscope objective (”underfilled aperture”). This illumination scheme ensures a well defined shape of the detection volume which can be described by a three-dimensional Gaussian volume paramterized by a radial (ω_0) and an axial (z_0) spatial extensions [11, 12]. Most relevant for the topic discussed here, the confocal setup allows to measure various fluorescence parameters from freely diffusing single molecules. In highly diluted samples (molecule concentrations in the order of a few ten picomolar) single fluorescently labled molecules cause ”bursts” of photons in recorded time traces which can be assinged to individual molecules. The approach of analyzing fluorescence parameters with single molecule sensitivity by employing a multiparamter burst analysis is described in Subsection 3.4.
In the case of surface immobilized molecules or larger objects the confocal setup can be used in a scanning mode. Technically this is for example realized by mounting the sample (or the objective) on high-resolution piezoelectric stages wich provide translation and fine focusing in all three spatial directions. This allows lateral 2D scanning of extended surfaces and provides a high resolution image of fluorescent surface tethered objects (see Fig. 4). The time needed to record a full image is however much larger (uo to minutes) as compared to TIRF microscopy (milliseconds to seconds). An additional scan along the optical axis gives even 3D information of the sample. Thus confocal laser scanning microscopy (CLSM) is a very important tool for *in vivo* sudies on 3D objects, like cells, with high resolution (see for example [13] and references therein).

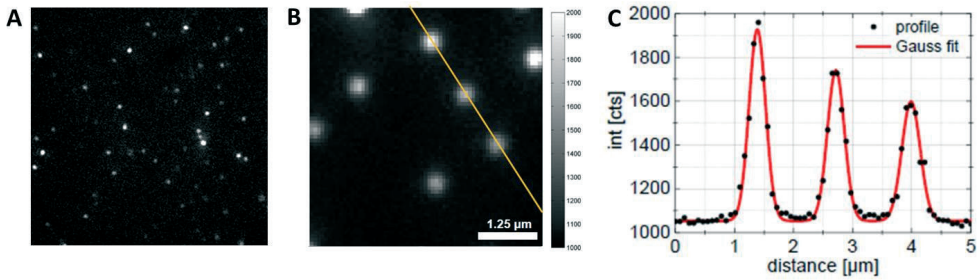


Fig. 4: *(A) Here surface immobilized fluorescent beads (100 nm in diameter) loaded with hundreds of fluorophores are imaged at high resolution. (B) The same sample as shown in (A) was imaged with a larger number of pixels per peak. With this magnification the resolution of the optical setup can be determined. (C) The result of a Gaussian fit of the intensity profile taken from measured counts along the straight line as shown in (B). The FWHM values of the Gaussian profiles (here approximately 500 nm) determines the resolution limit.*

3 Single-Molecule Data Acquisition and Analysis Methods

As discussed in the previous section, TIRF and confocal microscopy have advantages and disadvantages for applications in single molecule detection. In TIRF microscopy we can only study surface attached molecules with millisecond time resolution. But hundreds of molecules can be monitored at the same time over extended observation times, principally only limited

by the photostability of the fluorescent dyes. This allows to follow and to characterize many biological reactions in time, like catalytic reactions, protein unfolding and refolding transitions, or the movement of molecular motors [7, 14]. Confocal microscopy enables the detection of freely diffusion molecules, which is often a more physiological environment for biomolecules, but only with restricted observation times in the order of a few milliseconds. The later is determined by the dwell time of individual molecules in the detection volume. This exludes in many cases the observation of the full time course of the biological reaction. Therefore, the particular focus of the research topic determines in the end which of both intrumental approaches is best suited. In the following subsections some applications by employing different fluorescence parameter and different analysis approaches are described in more detail.

3.1 Förster Resonance Energy Transfer (FRET)

Förster Resonance Energy Transfer (FRET) is a very powerful approach to measure distances between two reference points to which fluorescent dyes are attached. FRET relies on non-radiative energy transfer between two fluorophores, termed donor and acceptor, and allows to measure distances and distance changes in the order of a few nanometer (Fig. 5).
The efficiency of the energy transfer $E(R_{DA})$ is a direct measure of the relative distance between both fluorophores R_{DA}:

$$E(R_{DA}) = \left[1 + \left(\frac{R_{DA}}{R_0}\right)^6\right]^{-1} \tag{1}$$

Here an important quantity, the Förster radius R_0, determines the regime of distances for which the measurement is sensitive (see Fig. 5D). This quantity depends on several properties of the employed dyes and on environmental conditions:

$$R_0[\mathring{A}] = 0.211 \cdot \left[J(\lambda) \cdot \kappa^2 \cdot \phi_D \cdot n^{-4}\right]^{\frac{1}{6}} \tag{2}$$

Here, $J(\lambda)$ is the overlap integral between the corrected fluorescence emission spectrum of the donor and the absorption spectrum of the acceptor (see Fig. 5C), κ^2 is the dye orientation factor set to $\frac{2}{3}$ for freely rotating dyes, ϕ_D is the quantum yield of the donor, and n is the refractive index of the solution (for more details see [15]). Experimentally we can determine the transfer efficiency values of *E*, either by measuring the donor and acceptor emission intensity (F_D, F_A, respectively)

$$E = \frac{F_A}{F_A + \gamma \cdot F_D} = 1 - \frac{\tau_{DA}}{\tau_{D0}} \tag{3}$$

where γ is a correction factor, or by measuring the donor lifetimes in the absence (τ_{D0}) and in the presence of the acceptor (τ_{DA}) (for more details see [15, 5]).
With respect to applications in life sciences, the obtained distance information can be employed to monitor association and dissocioation events in bi-molecular interactions or to follow conformational changes within the target molecules, if the dyes are attached in a proper manner (Fig.5A, 6A). Two examples of single molecule FRET studies are presented in the following dealing with conformational changes of proteins.

In the first case a hinge-bending protein is studied with respect to ligand induced conformational changes. The corresponding data were measured with freely diffusing double-labeled proteins

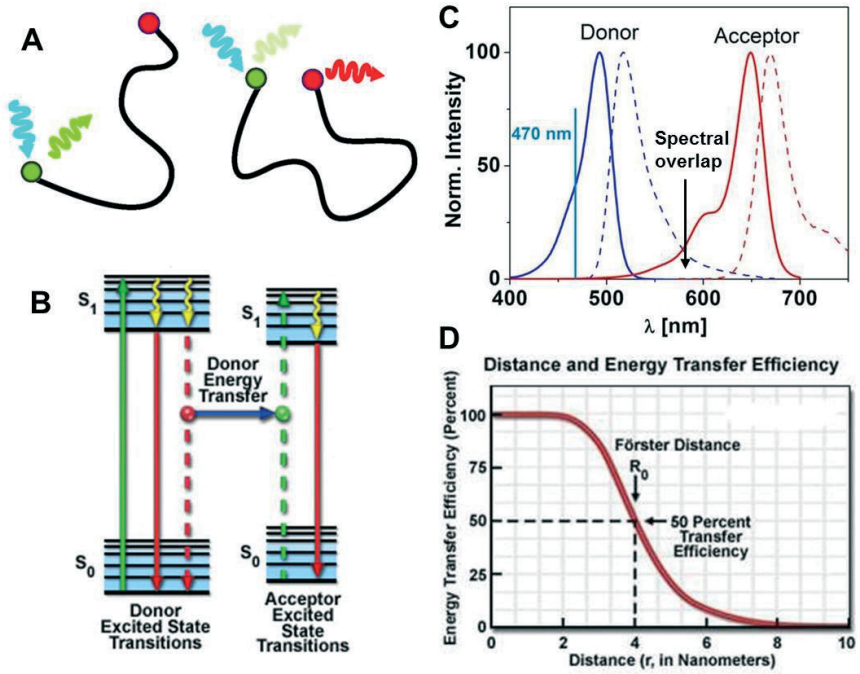


Fig. 5: *(A) For the molecule shown on the left, donor excitation leads only to donor emission, since the distance to the acceptor is to far (transfer efficiency $E = 0$), while the molecule on the right exhibits donor as well as acceptor emission, because the two fluorphores are closer to each other ($E > 0$). (B) Here a Jablonski diagram is illustrating the coupled transitions involved between the donor emission and acceptor absorbance in fluorescence resonance energy transfer. (C) Respective absorption and emission spectra are shown for a FRET pair (here for Alexa488 Alexa647 dyes). Based on this spectral information the overlap integral $J(\lambda)$ can be calculated. (D) In this graph the energy transfer efficiency is plotted as function of the inter-dye distance* r *or R_{DA}. Panel (A) from [19].*

in highly diluted solutions on a confocal microscope [16]. The investigated enzyme, Phosphoglycerate kinase (PGK), exhibits a pronounced hinge-bending movement upon ligand binding (Fig. 6A). In order to study this feature in more detail, distributions of conformational states and dynamical transition between the states were analysed by single molecule measurements. Typical raw data for those measurements are presented in Figure 6B, where fluorescence bursts for donor and acceptor channels are depicted in measured time traces. For each burst, which represents fluorescence emission from an individual molecule, the ratio of acceptor and donor counts allows to calculate a FRET efficiency (Eq. 3). Only bursts exceeding a threshold (e.g. 20-50 total counts for both channels per burst) are used for calculating FRET efficiencies. With measurements over several hours the related time traces provide thousands of efficiency values which can be histrogrammed (Fig. 6C).

By fitting the obtained histograms with Gaussian distributions valuable information about static and dynamic heterogeneity of the molecules in the sample can be extracted. (i) The histogram reveals the existence of two populations, one with a low FRET efficiency (E = 0.35) and one

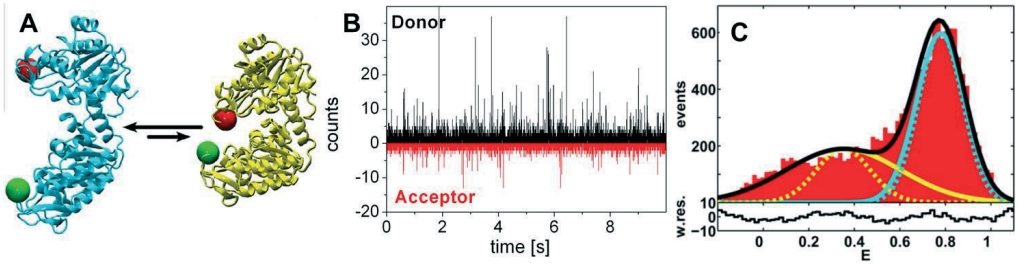


Fig. 6: *(A) For the investigatd enzyme the lable postions are highlighted by green and red spheres and a scheme of possible conformational changes demonstrates the pronounced inter-dye distance change caused by conformational transitions. (B) Time traces of the donor and the acceptor channel show regions of larger number of photon counts (bursts) within a certain time intervall (here a millisecond binning is used). (C) All bursts for which the FRET efficiency is calculated contribute to the given histogram. Here clearly two populations are visible, for which Gaussians were fitted (solid lines) to the distributions. From Ref. [16].*

with a high FRET efficiency (E = 0.8). According to equation (1) we can calculate the mean values of the respective inter-dye distances charcterizing the two conformational states (open state with with 56 $\mathring{A}$ and a closed state with 40.6 $\mathring{A}$). (ii) The area under each Gaussian is proportinal to the stastical weight (occupation probability) of the respective state (43 $\%$ for the open state and 57 $\%$ for the closed state). (iii) The width of the Gaussian can give information on a static distribution of states around the mean value (given by the centre of the Gaussian peak). Without any heterogeneity a shot-noise limited width is expected with $\sigma^2_{SN} = \frac{\langle E\rangle\cdot(1-\langle E\rangle)}{\langle N\rangle}$ where $\langle E\rangle$ is the mean efficiency and $\langle N\rangle$ is the average number of counts used for calculating the respective E-values (dotted lines in Fig. 6C, for details see [5]). Obviously we observe a pronounced additional broadening beyond the shot-noise limit only for the open state. This result indicates a static heterogeneity of the open state, where possible structural fluctuations of the protein structure may occur on a time scale slower than the dwell time of molecules within the detection volume (approximately a few milliseconds). Structural motions much faster than the dwell time average out during their transit through the detection volume and do not cause an additional broadening beyond the shot noise (see the population of the closed state). More detailed information on structural fluctuations of the proteins structure can be obtained by emplyoing correlation techniques (Section 3.2) or fluorescence lifetime measurements (Section 3.3).

In a further example of employing single molecule FRET surface tethered proteins were investigated by using TIRF microscopy. Although this approach requires a more elaborate sample preparation (Fig. 7A) the invaluable advantage is the possibility to monitor individual molecules over much longer times as compared to studies with diffusing molecules. This approach allows to track structural fluctuations within biomolecules over seconds (Fig. 7B) and sometimes even over minutes. In order to obtain rather long time traces, photostability issues have to be considered. A combination of a oxygen scavenging system and a triplet state quencher is in many cases the method of choice for enhancing photostability (for details see [18]). As shown in Fig. 7B donor as well as acceptor dyes provide fluorescence signals only over a few seconds until they bleach (here the acceptor after 3 seconds and the donor after 3.5 seconds). However, clearly two different individual states can be identified in the measured time traces, with a high

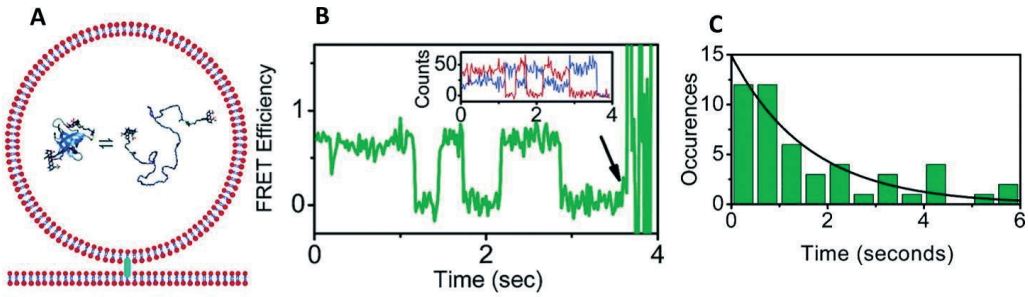


Fig. 7: *(A) Here a surface blocked cover glass slide is used to anchor a vesicle container with an entrapped single fluorescent labeled protein. (B) Based on the obtained intensity values from both channels (see lines in the inset; acceptor in red and donor in blue) the time course of the corresponding FRET efficiency can be calculated (green line). (C) Life periods of states measured from many single molecule time-traces are histrogramed. As a results a rate constant is obtained by fitting the decay in the histrogram. From Ref. [17].*

FRET state at E $\sim$ 0.8 and with a low FRET state at E $\sim$ 0. The most valuable information is given by the lifetimes of the states and the inter-conversion rates between states. This is unique information, achievable only from single molecule time traces with no need of sample synchronization. The investigation of more complex time traces with multiple states is best done by using statistical analysis techniques, like hidden Markov-moldelling or Maximum-likelihood approaches (see [7]). By histogramming the obtained life periods as measured with many time traces (Fig. 7C) we can also extract rate constants characterizing the underlying process, here folding/unfolding transitions at a certain concentration of a chemical denaturant.

3.2 Fluorescence Correlation Spectroscopy (FCS)

Fluorescence correlation spectroscopy (FCS) employs a statistical analysis on the time dependence of a fluctuating signal, typical a fluorescence intensity fluctuation. The fluctuations in this signal, for example originating from diffusing fluorescent molcules, can be employed in FCS to probe the process that cause them. For fluorescent molecules diffusing in the sample the effective concentration fluctuations of these molecules are monitored in an open detection volume as typically utilized in confocal microscopes (Fig. 8A).
If the average total number of fluorescent molecules which are present in the detection volume at the same time is rather low (e.g., 1-10 molecules) we observe a strong fluctuating signal for bright fluorophores (Fig. 8B). By employing a time domain auto-correlationg analysis we measure the self-similarity of a time series signal and the decay describes the temporal persistence of the information carried by it (Fig. 8C). The fluorescence correlation function of a fluctuating signal $F(t)$ can be written as (see for example [15, 2])

$$G(\tau) = \frac{\langle F(t)F(t+\tau)\rangle}{\langle F(t)\rangle^2} \tag{4}$$

where $F(t)$ and $F(t+\tau)$ are fluorescence intensity values at time t and $t+\tau$, respecively, while $\langle F(t)\rangle$ is the time average mean value of the signal (see blue points in Fig. 8C). The

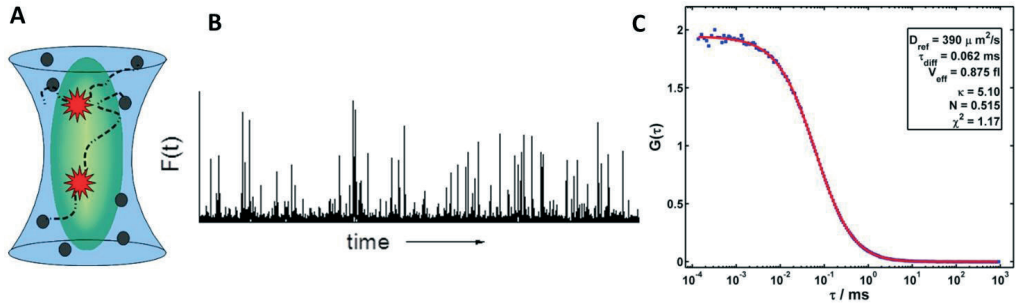


Fig. 8: *(A) The detection volume is shown together with fluorescent molecules which emit fluorescence photons when they travers the volume (red spheres) and stay dark when they reside outside the volume (black spheres). (B) The resulting fluctuations in the detected fluoescence intensity signal. (C) The auto-correlation of data shown in (B) gives a typical curve (blue points) which can be fitted by analytical model functions (red line) describing the underlying physical process.*

time τ is known as the corrrelation or lag time and is the time delay over which the fluctuations are compared. In order to extract physical parameters an analytical expression can be derived for fitting the experimentally obtained auto-correlation curve (using eq. 4). If we are dealing with (i) three-dimensional detection volumes that have a Gaussian shape and (ii) probing a translational Brownian diffusion process in 3D, the following expression can be used to analyse the data:

$$G(\tau) = \frac{1}{N}\left[1 + \frac{\tau}{\tau_D}\right]^{-1} \cdot \left[1 + \frac{\tau}{\kappa^2 \tau_D}\right]^{-\frac{1}{2}} \tag{5}$$

Here N is the average number of molecules in the detection volume and τ_D is the molecular diffusion time which is related by $\tau_D = \frac{\omega_0^2}{4D}$ to the translational diffusion coefficient D. The factor κ describes the eccentricity of the detection volume ($\kappa = \frac{z_0}{\omega_0}$, see Section 2.2). Obviously the auto-correlation function gives directly a measure of N for $\tau = 0$ with $G(0) = \frac{1}{N}$. One of the most valuable information that is achievable with FCS is the diffusion coefficient. By assuming spherical diffusing particles the Stokes-Einstein equation

$$D = \frac{k_B T}{6\pi\eta R_H} \tag{6}$$

delivers the hydrodynamic radius R_H which is a practical measure of the size of the diffusion molecule. A simple application of this approach is given in Figure 9 (AB) where a fluorescently labeled protein is measured in aqueous buffers with an increasing concentration of a chemical denaturant, called guanidine hydrochloride (GndHCl).

This chemical denaturant induces an unfolding of the protein which results in less compact protein structures and the apparent hydrodynamic radius increases (Fig. 9B). Therefore, the auto-correlation curves exhibit a shift towards larger diffusion times with increasing GndHCl concentrations (Fig. 9A). This shift is not only caused by a structural expansion of the protein induced by unfolding, but also by an increased solvent viscosity η which goes along with increasing denaturant concentrations in the solvent. To obtain meaningful results a correction for this effect is needed.

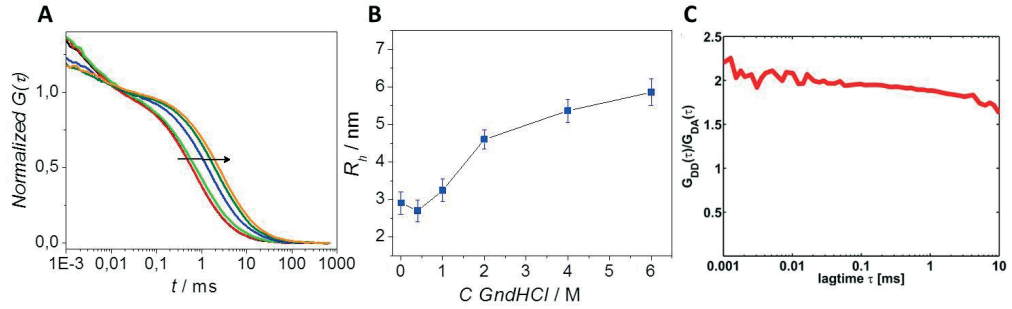


Fig. 9: *(A) Fits of auto-correlation curves are shown for PGK incubated in buffers with different GndHCl concentrations with an increase of τ_D for increasing GndHCl concentrations (shift of the curves as indicated by the arrow) (B) The resulting hydrodynamic radii (corrected for solvent viscosities η) are shown here as a function of the denaturant concentration. (C). For data presented in Fig. 6C the given ratio of the donor-donor auto-correlation and the donor-acceptor cross-correlation functions shows no decay in a time window up to 10 milliseconds (for details see [16]).*

Correlation analyses can also be employed for FRET data in order to investigate inter-conversion times for proteins in different conformational states (see Section 3.1, Figure 6C). For this purpose the more general variant of equation (4) given by

$$G_{ij}(\tau) = \frac{\langle I_i(t) I_j(t+\tau)\rangle}{\langle I_i(t)\rangle \langle I_j(t)\rangle} \tag{7}$$

is employed. This equation allows to calculate donor-donor auto-correlations $G_{DD}(\tau)$ and donor-acceptor cross-correlations $G_{DA}(\tau)$. Only burst selected photons were used for the analysis. From these two functions, the ratio $\frac{G_{DD}(\tau)}{G_{DA}(\tau)}$ was calculated which characterizes further fluctuations in time in addition to the apparent translational diffusion. As shown in Fig. 9C this ratio of the correlation curves is not decaying in a time window up to ten milliseconds. This result indicates that a transition between both conformational states (open and closed states) must be slower than 10 ms.

3.3 Fluorescence Lifetime and Anisotropy

In addition to studies based on intensity measurements the measure of fluorescence lifetimes is a futher valuable fluorescence parameter, which for example can be employed in the determination of Förster transfer efficiencies (see eq. 3). As already mentioned in Section 3.1, useful information about dynamic heterogeneity of conformational states is hidden in intensity based efficiency histograms if the related dynamics is faster than a few milliseconds (see Fig. 6C, closed state population with a mean E-value ~ 0.8). In this case fluoresence lifetime measurements can reveal information on the distribution width (i.e. amplitude of motion) around a mean inter-dye distance. Experimentally lifetime measurements in fluorescence spectroscopy are typically performed by employing pulsed excitation and time correlated single photon counting (TCSPC) which can be realized in a confocal microscope setup (for details see [19] and references therein). Using this appoach we obtain a fluorescence intensity decay in time $F(t)$ which

can be used to determine the fluorescence lifetime(s) τ_i by fitting $F(t)$ with a multi-exponetial decay

$$F(t) = I_0 \cdot \sum_i x_i \cdot e^{-\frac{t}{\tau_i}} \tag{8}$$

where I_0 is the fluorescence intinsity at $t = 0$. Caused by the fact that the lifetime of the donor fluorophor is reduced in the case of donor energy transfer to the acceptor (see Fig. 5B) the apparent donor lifetime τ_{DA} is given by

$$\frac{1}{\tau_{DA}} = k_F + k_{nr} + k_{ET} = \frac{1}{\tau_{D0}} + k_{ET} \tag{9}$$

with

$$k_{ET} = \frac{1}{\tau_{D0}} \left[\frac{R_0}{R_{DA}}\right]^6 \tag{10}$$

Since we are dealing with a distribution of inter-dye distances R_{DA} (related to dynamic heterogeneity of conformational states) the total decay $F_{DA}(t)$ is the sum of distance-dependent multi-exponential lifetime decays weighted by the probability $p(R_{DA})$ to find the dyes at distance R_{DA}:

$$F_{DA}(t) = \sum_i x_i \int_{R_{DA}} p(R_{DA}) \cdot e^{-\frac{t}{\tau_{D0,i}}\left[1+\left(\frac{R_0}{R_{DA}}\right)^6\right]} dR_{DA} \tag{11}$$

In this model, the physical information about the fluorescently labeled molecule is enclosed within the probability distribution $p(R_{DA})$, which depends on the portion of space accessible to the dyes and on their relative configuration. In general, it turns out that this distribution is well approximated by a Gaussian function:

$$p(R_{DA}) \; = \; \frac{1}{\sqrt{2\pi} \cdot \sigma_{DA}} \cdot e^{-\frac{(R_{DA}-\langle R_{DA}\rangle)^2}{2\cdot\sigma_{DA}^2}} \tag{12}$$

where the standard deviation σ_{DA} describes the average amplitude of the inter-dye distance fluctuations and the mean $\langle R_{DA}\rangle$ the most probable inter-dye distance. A comparison of intensity based single molcule FRET efficiencies, characterized by only one populated state (Fig. 10A), with a smFRET-filtered lifetime analysis of the same data (Fig.10B) is shown for a cases study with labeled doubled stranded DNA [20]. These rather rigid DNA molecules exhibit only fast motions caused by dye-linker flexibility (see below). One of the most fundamental advantages of using the lifetime analysis is given by the fact that the determination of the correction factor γ used in the ratiometric efficiency determination (see eq. 2) is not required, for which quantum yields of the donor and the acceptor as well as detection efficiencies for both colors must be known.

A further important fluorescence parameter is given by the fluorescence anisotropy. In order to measure this quantity, first, linear polarized excitation light is required (for example horizontally polarized) and second, one has to measure the vertical $I_\perp$ and the horizontal $I_\parallel$ component of the emitted fluorescence light. The latter can be realized by using a polarizing beam splitter cube in the emission light path (see Fig. 3A). The anisotropy r is given by [15]:

$$r = \frac{I_\parallel - I_\perp}{I_\parallel + 2I_\perp} \tag{13}$$

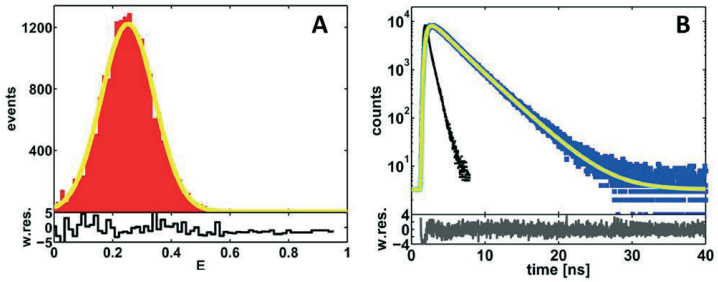


Fig. 10: *(A) A FRET efficiency histrogram of a double lableled double stranded DNA. The mean position of the peak maximum at $E \sim 0.25$ is related to a inter-dye distance of about 60 Å. (B) The corresponding lifetime analysis (using eqs. 11,12) yields a rather similar average inter-dye distance value and in addition a distribution width $\sigma_{DA} \sim 7.2$Å caused by fast dye motions. The latter value cannot be obtained from data shown in (A) since the width in the histogram shows no contribution beyond the shot noise limit. Taken from [20].*

In principle r provides a measure of the rotational mobility of the fluorescent dye which can be analysed by steady-state and by time-resolved measurements. With respect to FRET studies the rotational mobiliy of the dye is a crucial issue, because in most applications it is assumed that the employed fluorescent dyes are free to rotate while they are attached to the biomolecule. This assumption is manifested by the fact that κ^2 is set to a value of $\frac{2}{3}$ (see eq. 2), which originates from a dynamic isotropic averaging of the fluorophore orientations on a time scale faster than the inverse energy transfer rate $(\frac{1}{k_{ET}})$ [15]. The most straighforward way to prove this assumption is to measure time-resolved anisotropy decays $r(t)$ for both dyes. Corresponding results of such measurements are shown in Figure 11.

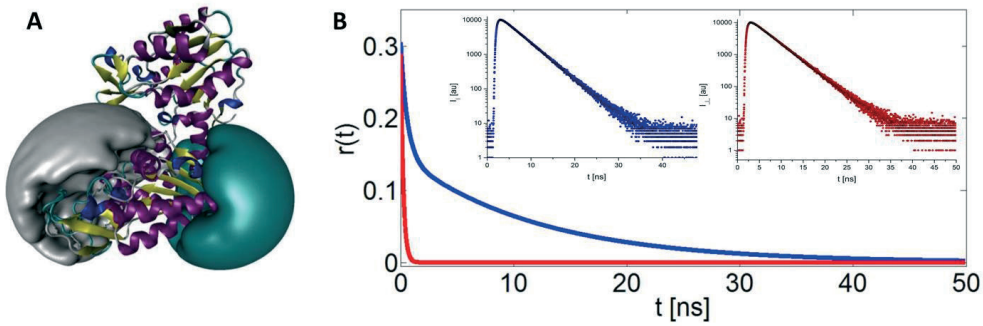


Fig. 11: *(A) Here a protein structure is depicted together with accessible volumes in which the attached dyes can rotate freely . (B) Time resolved anisotropy decays are shown for a freely diffusing dye in aqueous solution (red curve) and for the same dye attached to a specific position of the protein structure (blue curve). Here the free dye is characterized by $\theta_r = 0.2$ ns and the dye attached to the protein by $\theta_r = 0.6$ ns and $\theta_M = 12$ ns (inset: measured data from free dye for perpendicular and parallel components).*

Experimental lifetime decays for perpendicular ($I_\perp(t)$) and for parallell ($I_\parallel(t)$) components (see inset in Fig. 11B) are utilized to determine $r(t)$. According to the Perrin equation (see [15])

the measurement is sensitive for rotational motions on a similar times scale like the fluorescence lifetime of the employed dye (here $\tau \sim 4$ ns for Alexa 488). As a consequence for measurements with dyes attached to proteins we observe two different types of rotational motions occuring on different time scales: (i) a fast pure dye rotation enabled by a flexible linker tethering the dye to the protein and (ii) a slower overall rotation of the protein together with the dye, both characterized by corresponding rotational correlation times θ_r and θ_M, respectively. For this scenario a "wobbling-in-a-cone" model has proven to give a meaningful interpretation of the measured data:

$$r(t) = r_0 \cdot \left[(1 - A_1)e^{-\frac{t}{\theta_r}} + A_1\right] \cdot e^{-\frac{t}{\theta_M}} \tag{14}$$

where r_0 represents the fundamental anisotropy and A_1 an amplitude factor for restricted motions (for details see [21]). As visible from the obtained results of the fits given in Fig. 11B (see legend), the rotation of the dye attached to the protein, characterized by θ_r, is still fast enough to justify the assumption of $\kappa^2 = \frac{2}{3}$.

3.4 Multiparameter Burst Analysis

A very powerful approach in single molecule fluorescence studies aiming to characterize biomolecules with the highest structural and temporal resolution is to employ several fluorescence parameter in a joint manner. Some examples making use of this approach were already discussed in the previous sections. In general, one has access to a maximal number of relevant fluorescence parameters by employing pulsed excitation, with time correlated single photon counting, and a simultaneous detection of photons of different emission wavelengths and of different polarisations [22, 19].

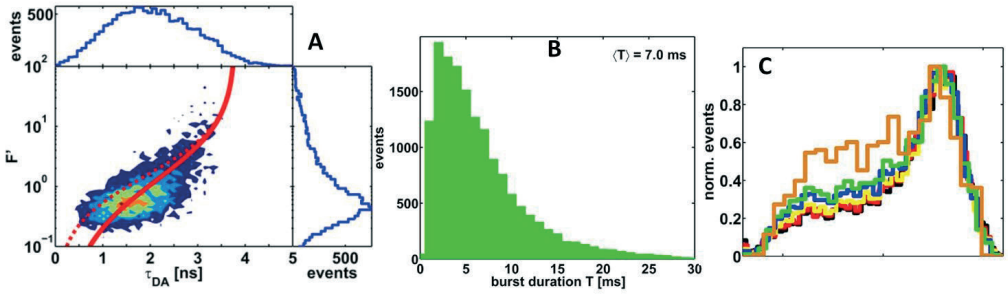


Fig. 12: *(A) Data presented already in Fig. 6C are shown here in a 2D-plot were in addition to intensity based efficiency parameters also donor lifetimes are included. The theoretical background of how theses plots can be utilized to identfiy fast dynamics present in the sample molecules is given in Ref. [19]. (B) The distribution of burst duration times is shown for a typical single molecule measurement with proteins diffusing in solution. (C) Here efficiency histograms are shown for data from selected bursts with different burst duration times (brown curve with T = 0.5 ms, the others vary from 1-10 ms). From Ref. [16].*

For this purpose often 4-6 detection channels are operated in parallel, implemented in setups similar to that described in Fig. 3A. In particular for measurements with diffusing molecules, a set of these fluorescence parameters can be obtained for each single burst of photons. In this

way the FRET efficiency for individual molecules cannot only be determined by measuring the photon counts in each detection channel (see Fig. 6AC), but also by measuring the fluorescence lifetime obtained from the same photons detected within the given burst (see eq. 3). An example of how this twofold determination of FRET efficiencies can by used to characterize dynamical aspects of the investigated protein is shown in Fig. 12A. These so-called 2D-plots allow to identify motions which are significantly faster than the dwell time of the analysed molecule in the detection volume. Only motions slower or in the regime of this dwell time will show populations falling on the static FRET line (broken red line in Fig. 12A). Obviously the data shown here originate from a sample which is characterized by much faster motions, indicated by the fact that the corresponding population lies below the static line. A more elaborated analysis employing simple models allows to chacterized this fast motion in more detail, for example in terms of amplitudes of motion (for details see [19, 16]).
A further important issue in multiparameter burst analyses is the determination of the burst duration time (or molecule dwell time). Since the individual trajectories of molecules passing through the detection volume can be quite different, the observed burst duration times exhibit a certain distribution (Fig. 12B). Although for the data presented here the average burst duration time is in the order of a few milliseconds with $\langle T \rangle = 7$ ms, some duration times are much larger, up to 30 ms. Dependening of the burst duration time the total number of photons per burst (N_T) can be lager or smaller, which ist also important for the statistical error of derived parameters (e.g. FRET efficiency). In addition T is also crucial for the visibility of dynamic transitions between two distinguishable states, which will either average out during the transit (if they are faster than T) or will be visible as two separated populations (if transitions are slower than T). In Fig. 12C histograms are shown for burst analyses which included data from bursts with different duration times (T = 0.5 - 10 ns) and thereby with different N_T. These data show for all burst duration times more or less the same distributions and populations of two conformational states (see for comparison 6C were all bursts were included). Only the data from the shortest burst duration time exhibit a visible deviation, due to the small corresponding N_T resulting in bad counting statistics.
In addition to already mentioned fluorescence parameters (e.g. intensities, lifetimes, ansiotropy values, quantum yields) which are measured also for different colors simultaneously, specific excitation schemes can by employed for characterizing the labeled molecules in the sample and to filter the bursts. By alternating laser excitation (ALEX) or pulsed interleaved excitation (PIE) donor and acceptor excitations are alternated in a time regime (below a millisecond) to allow sufficient time for both dyes to be excited during their transit through the detection volume [23, 24]. This scheme can be employed to monitor association and dissociation of potentially interacting molecules on single molcule level, each of them labeled with a different type of fluorophore. With respect to FRET studies this approach is rather useful in sorting out bursts originating from proteins which do not carry a fluorescent acceptor. The fraction of these molecules give often rise to a artifical low (or zero) FRET population, which can cause problems in data analysis.

4 Conclusion

The fluorescence based single molecule methods described here were mainly focused on thoses techniques which are already well established and used by a rather large community of applicants. However, single molecule methods are still evolving and branching out. Important

future developments deal with achieving smaller detection volumes, for example by employing nano-apertures (zero-mode waveguides [25]), whith improving photophysical properties of fluorophores [7], and with advances in labeling procedures for biomolecules [26].

5 Acknowledgement

I would like to acknowledge contributions from Matteo Gabba, Tina Züchner, Daryan Kempe, and Henning Höfig who performed most of the measurements and data analyses presented here.

References

[1] P. Tamarat, A. Maali, B. Lounis & M. Orrit, *J. Phys. Chem. A*, **104**, 1-16 (2000)

[2] C. Gell, D. Brockwell & A. Smith, *Handbook of Single Molecule Flourescence Spectroscopy*, Oxford University Press, Oxford (2006)

[3] A.N. Kapanidis & T. Strick, *Trends Biochem. Sci.*, **34(5)**, 234-243 (2009)

[4] B. Schuler, *ChemPhysChem*, **6**, 1206-1220 (2005)

[5] X. Michalet, S. Weiss & M. Jaeger, *Chem. Rev.*, **106**, 1785-1813 (2006)

[6] A.A. Deniz, S. Mukhopadhyay & E. Lemke, *J. R. Soc. Interface*, **5**, 15-45 (2008)

[7] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai & T. Ha, *Annu. Rev. Biochem.*, **77**, 51-76 (2008)

[8] J. Fitter, A, Katranidis, T. Rosenkranz, D. Atta, R. Schlesinger & G. Büldt, *Soft Matter*, **7**, 1254-1259 (2011)

[9] W.E. Moerner & D.P. Fromm, *Rev. Sci. Inst.*, **74**, 3597-3619 (2003)

[10] M. Boehmer & J. Enderlein, *ChemPhysChem*, **4**, 792-808 (2003)

[11] S.T. Hess & W.W. Webb, *Biophys. J.*, **83**, 2300-2317 (2002)

[12] J. Enderlein, I. Gregor, D. Patra & J. Fitter, *Curr. Pharm. Biotechnol.*, **5**, 155-161 (2004)

[13] G. Cox, *Materials Today*, **5**, 34-41 (2002)

[14] A. Yildiz in *Handbook of Single-Molecule Biophysics* , P. Hinterdorfer, A. van Oijen (Eds.), Springer Press, New York, 1-71 (2009)

[15] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy - Second Edition*, Kluwert Academic/Plenum Publ., New York, USA (1999)

[16] M. Gabba, S. Poblete, T. Rosenkranz, A. Katranidis, D. Kempe, T. Zuechner, R.G. Winkler, G. Gompper & J. Fitter, *Biophys. J.*, **107**, 1913-1923 (2014)

[17] E. Rhoades, M. Cohen, B. Schuler, and G. Haran, *J. Am. Chem. Soc.*, **126**, 14686- 14687 (2004)

[18] R. Roy, S. Hohng & T. Ha, *Nature Methods*, **5**, 507-516 (2008)

[19] E. Sismakis, A. Valeri, S. Kalin, R.J. Rothwell & C.A.M. Seidel, *Methods in Enzymology*, **475**, 455-514 (2010)

[20] H. Hoefig, M. Gabba, S. Poblete, D. Kempe & J. Fitter, *Molecules*, **11**, 19269-19291 (2014)

[21] P. Lamprou, D. Kempe, A. Katranidis, G. Büldt & J. Fitter, *ChemBioChem*, **15**, 977-985 (2014)

[22] J. Widengren, V. Kudryavtsev, M. Antonik, S. Berger, M. Gerken & C.A.M. Seidel, *Anal. Chem.*, **78**, 2039-2050 (2006)

[23] A.N. Kapanidis, N.K. Lee, T.A. Laurence, S. Doose, E. Margeat, S. Weiss, *Proc. Natl. Acad. Sci.*, **101**, 8936-8941 (2004)

[24] B.K. Müller, E. Zaychikov, C. Bräuchle, D.C. Lamb, *Biophys. J.*, **89**, 3508-3522 (2005)

[25] P. Zhu and H.G. Craighead, *Annu. Rev. Biophys.*, **41**, 269-293 (2012)

[26] A.T. Krueger and B. Imperiali, *ChemBioChem*, **14**, 788-799 (2013)

C 4 Single-Molecule Mechanics and Force Spectroscopy

V. Walhorn[1], T. Dierks, J. Mattay[2], N. Sewald[2] and D. Anselmetti[1]

[1]Faculty of Physics, Bielefeld University, Universitätsstr. 25
33615 Bielefeld

[2]Faculty of Chemistry, Bielefeld University, Universitätsstr. 25
33615 Bielefeld

Contents

1 Introduction & Motivation 2

2 Single Molecule Mechanics – General Concepts 2

2.1 Receptor-Ligand Dissociation Kinetics 3
2.2 Receptor-Ligand Interactions at the Single Molecule Level 6

2.2.1 Biomolecular Receptor-Ligand Interaction 6
2.2.2 Supramolecular Host-Guest Systems: Calixarenes 8
2.2.3 Host-Guest Systems: Photoswitchable Calixarene Systems 9
2.2.4 Counter Intuitive Receptor-Ligand Interaction - Catch Bonds 12

3 Summary and Outlook 14

References 16

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction & Motivation

How do molecules interact with each other? When Emil Fischer proposed his key and lock model in 1894 he introduced a completely new perspective on the interplay of molecules [1]. Briefly, as he investigated the catalytic activity of invertase and β-amylase towards different polysaccharides he found that both enzymes exclusively digest distinct polysaccharides. With virtually no knowledge about the protein 3D-structure he attributed these findings to the complementary shape of the reaction partners. Today we know that due to recognition forces these host guest systems specifically bind to each other. These interactions are governed by multiple comparable weak (and unspecific) interactions like hydrogen bonds, van-der-Waals and electrostatic forces or donor-acceptor interactions between complementary surfaces. However, in collaboration all these minute contributions sum up to highly specific and comparably strong bonds. Consequently, the functionality of many (bio-) chemical systems is governed by molecular recognition. Typical examples are the specific interaction of antibody and antigen, receptor and ligand, complementary DNA strands or supramolecular compounds that (self-) assemble to structures of higher complexity.

The understanding of the impact how (weak) molecular forces act on and between (single) molecules, is highly relevant for understanding and controlling the macroscopic properties and processes (for instance the immune response, (cellular) adhesion phenomena or the structural integrity of materials). In the last decade, single molecule force spectroscopy techniques have contributed a wealth of information on the impact of forces on the molecular scale that go beyond the common knowledge of macroscopic force experiments and hitherto have not been accessible. Upon stretching individual (bio-) polymer strands [2] as well as investigating and quantifying the interaction forces between macromolecules [3-5] a deep insight into elapsed force and binding mechanisms, their associated reaction pathways and reaction kinetics could be investigated. Furthermore, many experimental and theoretical verifications allowed to establish a coherent framework where the stochastic, thermally driven reactions at the single molecule level could be related to the macroscopic properties (observables) or thermodynamic state variables of a molecular ensemble (ergodic principle).

Within this survey, we briefly introduce general experimental concepts and the theoretical framework of single molecule force spectroscopy spectroscopy which is mandatory to analyze the data in a quantitative manner. The latter is fundamentally described by the Kramers-Bell-Evans theory (KBE-theory), which provides a coherent bridge between the nanoscopic force values determined in single-molecule experiments and the macroscopic ensemble parameters of the analyzed system (e.g. dissociation rate constants). Furthermore, we give a brief summary of certain publications to give an overview of contemporary SMFS applications.

2 Single Molecule Mechanics – General Concepts

Single molecule force spectroscopy (SMFS) has proven to significantly contribute in molecular binding studies to quantitatively determine:

- forces required to disrupt inter- and intramolecular (non)covalent bonds
- molecular elasticities (as mentioned before)
- dissociation rate constants (average lifetime of the complex)

- dimerization equilibrium constants
- details of the binding energy landscape
- entropic and enthalpic forces and energies required to stretch single molecules
- binding properties of "insoluble" molecules in solvent environment

in and between surface-bound but "unlabelled" molecules in an affinity range of 10^{-4}-10^{-15} M at the sensitivity of single point mutations or structural variations.

Here, we focus on typical intermolecular bond strengths (recognition forces) that are within the range of 40-200 pN and can nicely be accessed and investigated with several single molecule force spectroscopy techniques like atomic force microscopy (AFM), magnetic and optical tweezers, micropipettes etc. The experimental concepts presented here concentrate on AFM based SMFS. However, they commonly can be adapted to other SMFS techniques with ease.

In SMFS, two molecular binding partners of interest have to be immobilized and covalently bound to opposing surfaces (i.e. cantilever tip and sample surface). In order to first let the molecules associate, both surfaces will be approached to each other and brought into contact. Upon withdrawing the surfaces, a molecular binding event can be detected when the force transducer (cantilever) is withheld at the opposing surface. Repeated approaching and withdrawing (also termed as force cycle) result in force distance curves (Fig.2a). There, the force acting on the AFM tip is plotted against its vertical position (piezo extension), where a molecular dissociation event can be identified by a sudden jump in the "attractive" force regime back to the curve at zero force (F=0).

2.1 Receptor-Ligand Dissociation Kinetics

In thermal equilibrium non-covalently bound molecular complexes associate and dissociate at the same rate k_{on}^0 and k_{off}^0, respectively. This is easily expressed by the law of mass action

$$[L] + [R] \underset{k_{off}^0}{\overset{k_{on}^0}{\rightleftharpoons}} [LR],$$

where $[L]$ and $[R]$ are the concentrations of the ligand and receptor, and $[LR]$ denotes the concentration of the molecular complex. The complex dissociates thermally when it is driven across the activation energy barrier $\Delta G^{\#}$ (Fig.1). Correspondingly, the dissociation rate constant is expressed as follows:

$$k_{off}^0 = C\ exp\left(-\frac{\Delta G^{\#}}{k_B T}\right) \tag{1}$$

However, the pre-exponential factor C is commonly unknown and therefore k_{off}^0 cannot be estimated directly. The central concept in SMFS is to lower the activation energy $\Delta G^{\#}$ by applying an external force f and observe the force dependency of the molecular dissociation process. Correspondingly, the dissociation rate k_{off}(f) under external force can be expressed as

$$k_{off}(f) = C\ exp\left(-\frac{\Delta G^{\#} - f x_{\beta}}{k_B T}\right) \tag{2}$$

$$\Leftrightarrow k_{off}^{0}\ exp\left(\frac{f x_{\beta}}{k_B T}\right) \tag{3}$$

where x_β is the so-called interaction length (also termed as reaction length).

Generally, one can discriminate between two experimental techniques: In the force ramp approach the molecular complex is loaded at an arbitrary but constant velocity (f becomes a function of time) and the dissociation forces F_{diss} are acquired (Fig.2a). Typically, hundreds to thousands individual force curves have to be measured and combined in a force histogram (Fig.2a inset). These show a certain scatter since the process of molecular dissociation is of stochastic nature. The most probable dissociation force F_{max} can be determined by fitting an appropriate distribution function (like a Gaussian) to the measured force histogram.

The first theoretical description has been first published by Evans and Ritchie [6] who applied the general picture of the Kramers-rate based Bell adhesion model to this nanobiological dissociation problem with a ramped external force [7, 8]. Since the velocity of the force transducer in this active force-induced dissociation process is relatively slow (compared to the Brownian motion of the molecules) it can be described as a thermally activated decay with a force-dependent activation energy.

Within this model the theory predicted an increase of the observed dissociation forces with increasing loading rate, as it could be verified in many single molecule experiments in the last decade.

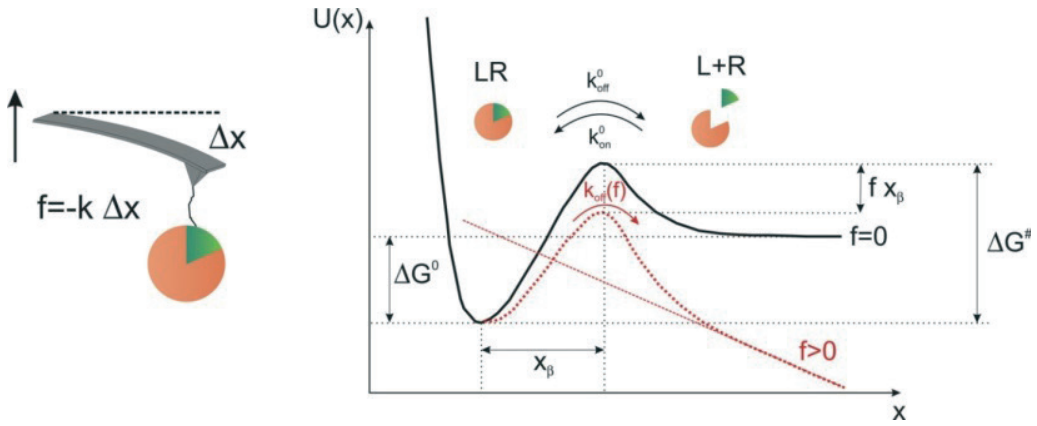


Fig. 1: *General concepts of SMFS. Left, a force f applied on a molecular complex can be estimated easily by the product of the cantilever spring constant k and the cantilever deflection* $\boldsymbol{\Delta x}$*. Right, molecular binding potential without (black) and with (dotted red) external force. By applying a force the activation energy barrier* $\boldsymbol{\Delta G^{\#}}$ *is lowered by* $\boldsymbol{f\ x_\beta}$*.*

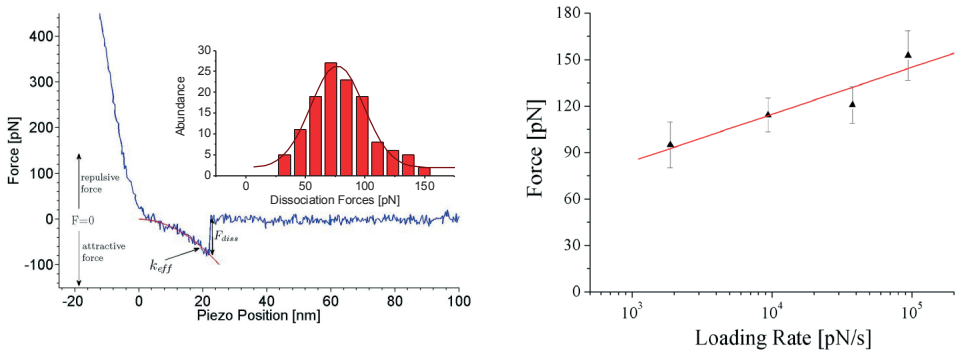


Fig. 2: *a) Typical force-distance curve (only retract curve) exposing a single molecule dissociation event. The dissociation Force F_{diss} is equal to the step height of the graph whereas the effective spring constant k_{eff} of the cantilever linker system is approximated by estimating the slope at the point of dissociation. The unbinding forces of several dissociation events are plotted in a histogram and the most probable dissociation force F_{max} can be estimated by approximation of a suitable distribution (inset). b) Semi-logarithmic scatter plot of F_{max} versus the loading rate (pulling velocity times effective spring constant). Equation 4 is approximated to the data to estimate $\boldsymbol{k_{off}^0}$ and $\boldsymbol{x_\beta}$(red plot).*

The most probable dissociation force F_{max} can be estimated with ease by:

$$F_{max} = \frac{k_B T}{x_\beta} \ln \frac{x_\beta \, r}{k_B \; T \; k_{off}^0} \tag{4}$$

F_{max} obviously depends logarithmically on the loading rate r which is the product of the experimental ramp speed v and the molecular elasticity k_{eff} (effective spring constant). The latter can be estimated as a linear fit to the force-distance curve just before dissociation (Fig.2a). The parameters x_β and k_{off} can now be extracted by fitting the experimental $F_{max} - r$ data in a semi-logarithmic representation (Fig.2b). Whereas the reaction coordinate x_β represents the distance between the minimum of the molecular binding potential and the transition state (activation energy) and can therefore be interpreted as the depth of the binding pocket (energy landscape, see Fig.1), the off-rate constant is defined as a kinetic parameter in thermal equilibrium of a molecular ensemble (mass action law). From a physical point of view, $k_{off}^{-1} = \tau$ is the average lifetime of the complex and gives information on the stability of the complex.

In force clamp SMFS a molecular complex is probed with a constant external force f until it dissociates (Fig.3a). Likewise, several single molecule dissociation events (individual life-times) have to be acquired. According to $N(t) = N_0 \; exp\left(-\frac{t}{\tau}\right)$ the average life-time τ for a given force is estimated by logarithmically plotting the number of intact bonds $N(t)$ versus time and approximating the (negative) linear slope to the dataset (Fig.3b). This procedure is repeated for several loading forces. The resulting dataset can be approximated by the inverse of equation 3 to estimate the complex life-time in thermal equilibrium τ ($k_{off}^{-1} = \tau$) and the reaction coordinate x_β (Fig.3c). Evidently, both approaches are equivalent as they both yield

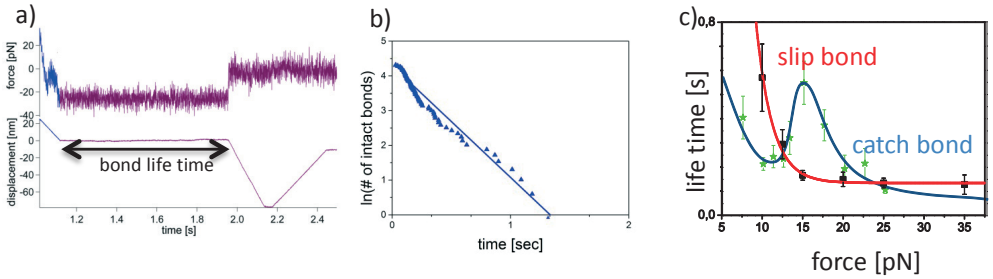


Fig. 3: *a) Force clamp dataset. The force acting on the cantilever and the piezo displacement are plotted versus time. The marked constant regime can be attributed to the life time of a molecular complex. b) N(t) plotted versus time for several individual life time measurements. The average complex life time is approximated by the slope of the plots. c) The average life times for several clamp forces (scatter plot) are approximated by equation 3 in the case of (common) slip bond dissociation (red graph) or by an appropriate catch bond model which will be introduced in chapter 1.3.4.*

the complex life time τ and the reaction coordinate x_β. However, dynamic force spectroscopy is much more common as the experiments can be conducted faster and the experimental setup does not need a precise and robust feedback loop for the load control.

Unfortunately, independent of the experimental approach the on-rate constant k_{on} cannot simultaneously be accessed, which makes a determination of the dissociation constant (constant of equilibrium) $K_d = k_{off}/k_{on}$ or the binding energy $E_b = k_BT \ln K_d$ difficult. However, in the limit of a diffusion-controlled reaction, an estimate of $k_{on} = k_{smol}$ via the Smoluchowski theory is possible, allowing for a quantitative approach to determine K_d and therefore an affinity ranking of the molecules involved [9]. Here, $k_{smol} = 4\pi R D_{SE} N_A$, where R, D_{SE}, N_A denote interaction distance, diffusion coefficient (Stokes-Einstein: $D_{SE} = k_BT/6 \pi \eta r_{prot}$) and Avogadro number, respectively.

2.2 Receptor-Ligand Interactions at the Single Molecule Level

2.2.1 Biomolecular Receptor-Ligand Interaction

Molecular recognition between receptors and their related ligands govern countless processes in biological systems such as immune response, enzymatic activity, signal transduction or genome replication. Despite the detailed knowledge about the structure and function of the corresponding receptor-ligand complexes, information about their association and dissociation kinetics is often lacking. Apart from ensemble techniques like surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC) SMFS can be a powerful means to explore host guest interactions especially when the availability of the compounds is limited due to low yield synthesis/expression or if the binding partners expose a low affinity (dissociation constant K_D in the micromolar range).

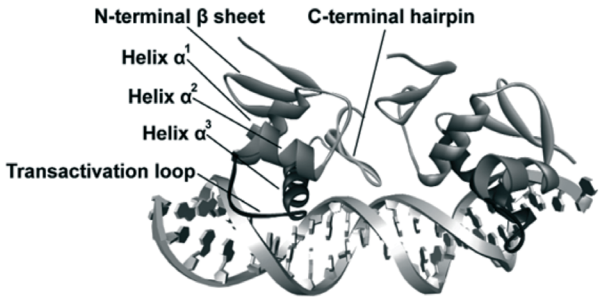


Fig. 4: *Crystal structure of two PhoB^{DBD}-proteins bound to a pho box in a head-to-tail arrangement. Figures adapted with courtesy [10].*

The transcription factor PhoB is a protein that is part of a two-component signal transduction system. Briefly, the transmembrane protein PhoR activates PhoB by phosphorylation of Asp53 in the regulatory domain and the concomitant structural change enables PhoB to bind to its cognate Pho boxes in the promoter region [10, 11]. This causes an alteration of DNA structure in the complex that enables interaction of the RNA polymerase with PhoB and subsequently with DNA, thus activating the transcription and biosynthesis of proteins. Notable, at least 287 genes are differentially expressed in the presence of active PhoB.

Here, the recognition and binding is a cooperative process, that is mediated multiple structural features. The DNA binding domain (DBD) of PhoB comprises a so-called winged helix-turn-helix recognition (wHTH) motif. An alpha helical recognition helix (α^3) mainly interacts with the major groove of the cognate DNA, a second helix (α^2) stabilizes the protein DNA complex, and a c-terminal beta hairpin penetrates the minor groove of the DNA recognition sequence (Fig.4). In order to analyze the recognition and to elucidate the minimal requirement for this cooperative binding process Wollschläger et al. analyzed the affinity of the whole DBD of PhoB (comprising the complete wHTH) and the α^3-recognition helix alone towards the cognate DNA sequence, respectively (Fig.4). Furthermore they introduced point mutations at selected loci in the α^3-recognition helix and observed the impact on the binding properties. The DNA–protein complex dissociates with k_{off}=0.0025 s^{-1}, which corresponds to an average life time of t=400 s. In comparison, the results obtained with the wild-type the α^3recognition helix alone yielded k_{off}= 3.1s^{-1}, thus indicating that the protein complex with the entire DBD

		k_{off} [s^{-1}] peptide (A) PhoB (190-209)	τ [s]		k_{off} [s^{-1}] protein (B) PhoB (127-229)	τ [s]
◆ wildtype	(1)	3.1 ± 2.1	0.32	(5)	0.0025 ± 0.0021	400
■ R193A	(2)	0.071 ± 0.053	14	(6)	0.012 ± 0.008	83
▲ H198A	(3)	50 ± 21	0.020	(7)	0.76 ± 0.25	1.3
R203A	(4)	—	—	(8)	—	—

Fig. 5: *AFM–DFS of PhoB peptides and proteins. A) PhoB peptides 190–209. B) PhoB DBD 127–229. k_{off}: dissociation rate constant for the peptide/protein DNA complexes; $\tau = {}^{1}/_{k_{off}}$: complex lifetime. The kinetic data were measured in 100mM Na2HPO4/50mM NaCl (pH 7.4). Figure adapted with courtesy from [11].*

dissociates about 1000 times more slowly, most likely because the full DBD incorporates additional amino acid residues that support complex stabilization.

Things are different when comparing the lifetimes of the mutants R193A and H198A on the protein and peptide level (R203A exhibited no binding). In both mutants the full protein sequences, i.e. the DNA binding domain PhoB (127-229), yield a distinctly longer lifetime than the mutated PhoB(190–209) peptide sequences. For protein mutant R193A (6) the lifetime is 83s, and the result for peptide mutant R193A (2) is 14 s (Fig.5). As shown by the work of Wollschläger et al., SMFS can reveal subtle differences between single point mutants that could not be detected to this extent with other methods. The results are of special interest for the future design of synthetic peptide and protein ligands as artificial DNA binders and transcription factors in synthetic biology.

2.2.2 Supramolecular Host-Guest Systems: Calixarenes

Supramolecular chemistry, which has been defined as "*the chemistry beyond the molecule*" [12, 13] deals with the chemistry and collective behavior of organized ensembles of molecules [14] where the rational and synthetic fabrication of molecular structures of increasing size, complexity and functionality is possible. Within this concept calix [4] arenes are versatile building blocks in supramolecular chemistry because of their three-dimensional structure and the various functionalizations that can be introduced at different positions of the molecular framework. Their ability to adopt chalice-like conformation makes them ideally suited as the basis for molecular receptors (Fig.6a,b).

Furthermore, well-defined self-assembled monolayers on gold can also be obtained if thioether moieties are introduced at the part termed the "lower rim" of the molecule.

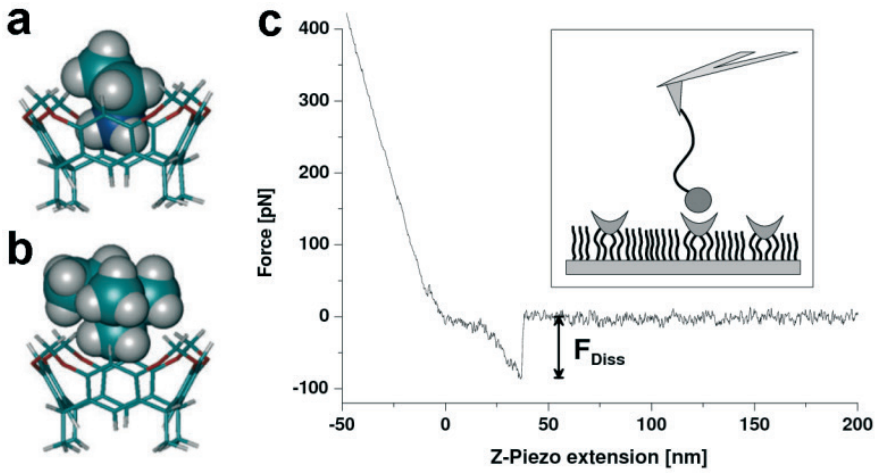


Fig. 6: *Supramolecular host-guest complex structure with calixarene cavitand and (a) an ethyl ammonium ion and (b) an ethyl trimethyl ammonium ion. (c) Typical single molecule force-distance curve of complex a) exhibiting an unbinding or dissociation force of ~85 pN. Inset: experimental scheme. Figures adapted with courtesy from [25a].*

In 2005 at Bielefeld University, the first single molecule study of supramolecular complex formation in a calixarene system was published, where the data could be quantitatively analyzed in accordance with the theoretical KBE standard model – a result that was in contrast to the one described above. Here, ammonium derivatives as the guest ligands were measured against resorcin[4]arenes cavitands in ethanol at room temperature using commercial instrument that was equipped with custom-made control and data acquisition system [15,16] [25a,d]. The 2,8,14,20-tetra-(10-(decylthio)decyl) cavitands were immobilized in diluted cavitand monolayers in a 1:40 mixture with didecylsulfide. The guest cations (ammonium (**A**), trimethyl ammonium (**TMA**) and triethyl ammonium (**TEA**), carrying one chemically modified entity each) were covalently attached to the AFM tip via a flexible polyethylene glycol (**PEG**) linker.

In AFM-SMFS experiments molecular dissociation events could be identified (Fig.6c) whereas their specificity could be verified in competition experiments (Fig.7a).
Interestingly, only the two smallest guests **A** and **TMA** were able to specifically bind to the cavitand, whereas the large **TEA** cation with a calculated diameter of 0.8 nm was sterically not able to enter the 0.7 nm wide host cavity.

Loading rate dependent dynamic force spectroscopy experiments and quantitative binding analysis according to the KBE-theory yielded $k_{off} = (0.99 \pm 0.81)\ s^{-1}$ for **A** and $k_{off} = (1.87 \pm 0.75)\times10^{-2}\ s^{-1}$ for the **TMA** residue, resulting in a bond lifetime of $\tau = 1.01$ s and $\tau = 53.5$ s, respectively. This finding, together with the results of the competition experiments, indicates that the **TMA** residue fits tighter into the receptor cavity, ensuring a rise in binding affinity as compared to **A**.

Upon assuming a diffusion limited association with a typical on-rate constant of $k_{on} = 10^5\ M^{-1}s^{-1}$, equilibrium constants for these reactions of $K_{diss} = 0.99\ s^{-1}/10^5\ M^1s^{-1} = 10^{-5}$ M (corresponding to $\Delta G = -28$ kJ mol^{-1}) and of $K_{diss} = 2\times10^{-2}\ s^{-1}/10^5\ M^{-1}s^{-1} = 2\times10^{-7}$ M (corresponding to $\Delta G = -38$ kJ mol^{-1}) for **A** and **TMA** ions, respectively, can be deduced. From the inverse slope of the loading rate dependency (Fig.7b) the molecular reaction length can be extracted yielding $x_\beta = (0.22 \pm 0.04)$ nm for **A**, and $x_\beta = (0.38 \pm 0.06)$ nm for the **TMA** ions. These estimated values qualitatively scale with calculated van-der-Waals diameters of 0.3 nm for **A** and 0.6 nm for **TMA**, respectively.

Therefore we can conclude that hydrogen bonds (not present in the **TMA** cavitand interaction) and cation-π-interactions have to contribute considerably to the molecular binding mechanism. Especially, the **TMA** cavitand system may profit from the latter contribution due to its positive charge distribution, which is shown to reside on the hydrogen atoms of the methyl groups.

2.2.3 Host-Guest Systems: Photoswitchable Calixarene Systems

The introduction of additional functionality and external control to mesoscopic systems is fascinating by itself, however, also aims for the rational design and directed synthesis of supramolecules that can mimic the function of biomolecules in distinct biomedical and technical applications. With this reverse engineering approach artificial and robust molecular motors and synthetic machines can be anticipated. Since future nanomachines rely on cyclic operation and external control, repetitive transition between at least two different molecular states e.g. by the conformational transition (switching) between two structural isomers is mandatory.

a) b)

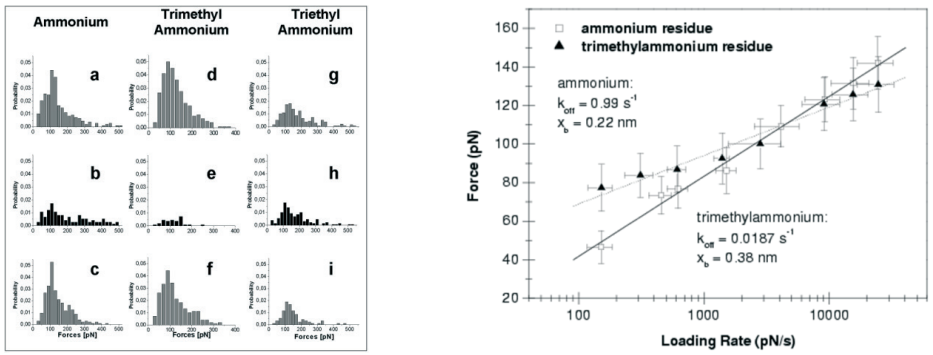


Fig. 7: *a) Activity status of probed cavitand surfaces against the three different ligands. (a,d,g) Single molecule force histograms measured in ethanol; (b,e,h) Control experiment: Single molecule force histograms in ethanol saturated with free competitive ions; (c,f) Reactivating the surface: single molecule force histograms after washing with pure ethanol restored the original unbindig probability. (Whereas the ammonium and trimethyl ammonium ligands exhibited specific interaction, only unspecific binding could be verified with the triethylammonium ligand (g,h,i)). b) Dynamic force spectroscopy. The loading rate dependent dissociation forces are logarithmically plotted for the binding of the ammonium and trimethyl ammonium guest residues to the resorc[4]arene cavitand. Figure adapted with courtesy from [16].*

As external control mechanism and transition stimulus the interaction of molecules with light, electric, chemical or mechanical potentials are known. External activation via electronic excitation, energy transfer, and/or ionic transport by light harvesting complexes or energy up conversion in photosynthesis is well known.

Fundamental for all these processes is the ability to convert electromagnetic energy into conformational changes where specific and noncovalent bonds can be formed and released due to affinity changes. In order to investigate such phenomena in an artificial model system, we have synthesized a bistable supramolecular host-guest system where the supramolecular receptor cavity of the aforementioned resorcin [4] arene cavitand has been combined with two photodimerizable anthracene moieties whose structures can externally switched by UV-light and temperature [15, 17, 18].

Since the photodimerization process blocks the entrance of the cavitand, the affinity properties of this photochemical macrocycle was investigated in AFM-SMFS experiments against guest cation complexation. (For scheme see Fig.8a). The transition from the open (active) to the closed (inactive) structure of this photochemical single-molecule switch and vice versa was performed by irradiation with a UV lamp at (368 ± 7) nm for 5 min and locally heating the sample to 60°C for 2 h, respectively.

Whereas the macroscopic switching properties were analyzed and verified in UV absorption experiments (see Fig.8b) the nanoscopic affinity modulation of this optical switch was investigated by AFM-SMFS, where the photoactive cavitand was immobilized on a gold

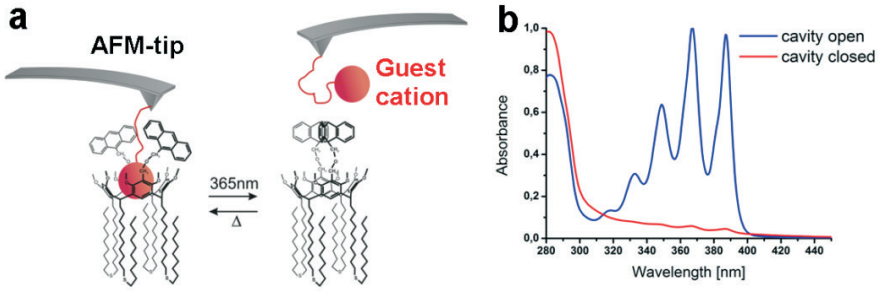


Fig. 8: *Resorcin[4]arene photoswitch. (a) Pictogram visualizing the photoactive cavitand receptor that is affinity probed against its ammonium cation ligand by AFM-SMFS. The switching between the open (active) and closed (inactive) structure is externally controlled by UV-light and thermal energy and mimics biological regulation triggers. (b) Resorc[4]arene photoswitch UV absorption spectrum for the open and closed isomer. The observed peaks for the open isomer are characteristic for the anthryl groups. Figures adapted with courtesy from [18].*

surface in a SAM in a 1:40 dilution with di-n-decyl sulfide. The guest molecule, an ammonium ion, was immobilized via a PEG-linker to the AFM tip.

Five series of AFM-SMFS experiments were performed in ethanol and presented as force histograms in Fig.9. As we have seen before, force histograms can be used as activity monitor. Fig.9a-d summarize the consecutive single molecule activity of the photoactive cavitand against ammonium complexation during the following stages: (a) open isomer after heating, (b) closed configuration after UV exposure, (c) (re)opened isomer after heating, (d) open isomer blocked with free ammonium and (e – image not shown) open isomer with full activity after washing with ethanol. These experiments prove that the photoactive resorc[4]arene cavitand can be reversibly switched between two different isomers that can be affinity probed by AFM-SMFS. Whereas one of the two isomers shows a high affinity to ammonium, the other exhibits almost none. Interestingly, the transition between the two isomers and its change in complexation affinity is of course related to a change in 3D structure of this supramolecular macrocycle.

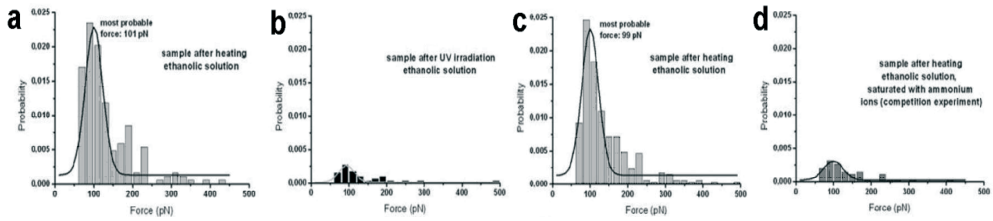


Fig. 9: *Single molecule force histograms of the resorcin[4]arene photoswitch. (a) Experimental series after heating the sample to 60°C for 2 h (open isomer and active state). (b) Series after irradiating the sample at 368 nm for 5 min (closed - inactive). (c) Series after renewed heating the sample for 2 h (open - active). (d) Series of competition experiments, performed in ethanol solution saturated with free ammonium (open/blocked - inactive). All experiments were performed with the same AFM tip and cavitand surface at a retract velocity of 1000 nm s^{-1}. Figures with courtesy from [18].*

2.2.4 Counter Intuitive Receptor-Ligand Interaction - Catch Bonds

As denoted before, noncovalent biological adhesion involves a multitude of different aspects like binding affinities, selectivities, multidomain interaction, force and/or allosteric regulation and many different molecular materials. Biophysically, we distinguish between slip and catch bonds. Whereas slip bonds are weakened, catch bonds are strengthened by tensile mechanical forces, respectively. Slip bonds were originally introduced by Bell in 1978 and mathematically treated within the framework of chemical reaction rate theory [7]. Catch bonds were conceptually introduced in 1988 by Dembo and co-workers [19]. Their experimental proves are tightly connected to the biological leukocyte recruitment system P-selectin (PSel) and its P-selectin glycoprotein ligand 1 (PSGL-1) [20]. Although, in the meantime numerous examples of slip bond-like interactions [4, 11, 15, 16, 21-29] could be demonstrated only a handful of catch bond systems could be identified in cellular and molecular force assays [30-34].

As mentioned in the Concepts section force ramp and force clamp SMFS yield the same bond parameters. However catch bonds are most commonly explored by the latter as the prolonged life time can be directly discerned from the resulting force-life time plot.

Beyond its experimental findings a couple of theoretical models were formulated helping to rationalize this counter-intuitive phenomenon. Here, we will briefly sketch two common models: The first approach was to introduce an alternative dissociation pathway along which the system can dissociate against low external forces resulting in a tightening of the bond [35-37]. At a certain critical force the system reaches its maximum stability and switches to the slip dissociation path. This approach is commonly referred to as one state two path model. The dependence of the dissociation rate k(f) on the applied force is given by

$$k(f) = k_c \exp\left(\frac{f\,x_c}{k_B T}\right) + k_s \exp\left(\frac{f\,x_s}{k_B T}\right) \tag{5}$$

Here, k_s and k_c are the dissociation rate constants for the slip (s) and the catch (c) dissociation pathway, x_s and x_c are the widths of the corresponding energy barriers and $k_B T$ is the thermal energy. Especially the catch bond characteristics of P-Selectin/PSGL-1 has been approximated reasonably well by this approach [36-38]. Of note, this model implies brittle bonds at zero load as the complex lifetime decreases for diminishing external forces.
However, there are host guest systems like hydrophilic domain (HD) of the human extracellular enzyme sulfatase-1 (Sulf1) and glycosaminoglycan heparansulfate (HS) that expose catch bond behavior in an intermediate force range (approx. 10 – 20 pN) and slip dissociation for vanishing external force (Fig.10a).

Correspondingly, a one state two path approach is not applicable to systems of this type as it diverges in the low force regime. More elaborate approaches have been reported that take account of force induced deformations [39], protein water interfaces [40], fluctuating energy barriers [41] or two bound states separated by an energy barrier [31, 35, 42-47] which we will present here:

By introducing a coupled, double-well energy landscape with two well confined binding states S_1 and S_2 separated by the energy barriers E_{12} and E_{21}, respectively (Fig.10b).

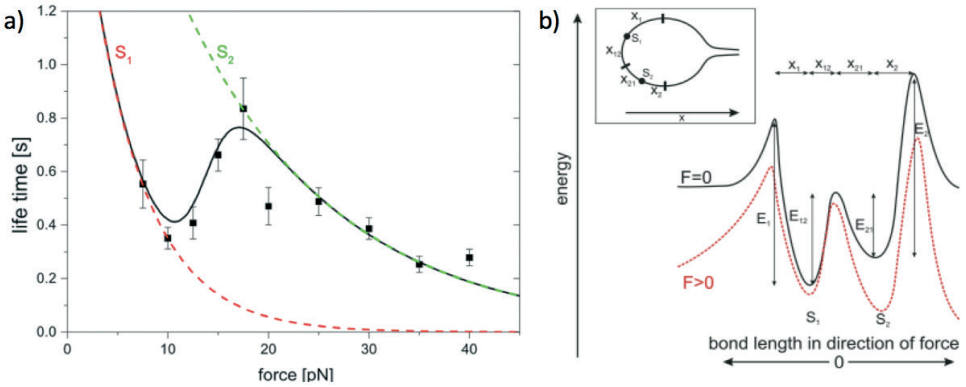


Fig. 10: *a) Experimental complex lifetimes for the HD substrate HS under a constant external force in the range of 7.5 – 40 pN (scatter plots). HS shows explicit catch bond behavior in the force range of 10 – 18 pN. The dashed red and green plots represent the slip dissociation from the individual binding states S_1 and S_2, respectively. b) Flat and 2D (inset) representation of the proposed energy landscape for HD/HS interaction. The occupancy of both bound states S_1 and S_2 is governed by equilibrium thermodynamics. In the force free state (black plot) solely S_1 is populated and dissociation can be observed from this state only. Upon applying a force the binding potential is tilted (red plot). In the intermediate force (transition) regime the system can flip between S_1 and S_2 by surpassing the internal energy barriers E_{12} and E_{21}, respectively. Consequently, the observed unbinding events are a superposition of both states. Increasing the force further depopulates S_1 successively and only dissociation from S_2 can be observed. Figures adapted with courtesy from [31].*

Both states individually obey slip bond characteristics and dissociation can occur from either of the two states depending on the external force. The total width of this barrier is given by $x = x_{12} + x_{21}$. The system can dissociate either from the low force state S_1 via x_1 or from the high force state S_2 via x_2 by crossing the corresponding barriers E_1 or E_2, respectively.

We can assume that both binding states can be attributed to different molecular conformations. Hence, it is reasonable to neglect higher order dissociation (e.g. S_1 via x_{12} and x_2) or dissociation from transition states when the conformational relaxation is fast compared to a single pulling experiment. As protein folding dynamics are in the range of micro to milliseconds [48-50] the upper temporal limit for HD conformational dynamics can be estimated to be in this range or even below. For comparison, the timescale of a molecular stretching experiment is in the range of seconds to tens of milliseconds.

Consequently, the population of the states S_1 and S_2 is in thermodynamic equilibrium at any time of the experiment and can therefore be calculated by equilibrium thermodynamics. To estimate the (force dependent) population of the two states we introduce the canonical partition function $Z(f)$ as a function of the external force.

$$Z(f) = \exp\left(\frac{E_{12} - f\,x_{12}}{k_B T}\right) + \exp\left(\frac{E_{21} - f\,x_{21}}{k_B T}\right) \tag{6}$$

With the energy difference $\Delta E = E_{12} - E_{21}$ and the compliance length $\Delta x = x_{12} - x_{21}$ of both states, we derive the population probability $p_1(f)$ and $p_2(f)$ of the states S_1 and S_2.

$$p_1(f) = \left(1 + \exp\left(-\frac{(\Delta E - f\,\Delta x)}{k_B T}\right)\right)^{-1} \quad (7)$$

$$p_2(f) = \left(1 + \exp\left(\frac{(\Delta E - f\,\Delta x)}{k_B T}\right)\right)^{-1} \quad (8)$$

In line with the afore-introduced Kramers-Bell-Evans (KBE) model for slip bonds ====[6-7] we define the total dissociation rate $k(f)$ as the probability weighted sum of the dissociation rates from S_1 and S_2.

$$k(f) = p_1(f) k_1 \exp\left(\frac{f x_1}{k_B T}\right) + p_2(f) k_2 \exp\left(\frac{f x_2}{k_B T}\right) \quad (9)$$

Here, the state from which the system dissociates explicitly depends on the applied force f, the shape of the binding potential landscape, parameterized by the compliance length Δx and the energy difference ΔE between S_1 and S_2.

Within this theoretical framework one explicitly observes three different dissociation regimes at low, medium and high forces: At low and high forces solely S_1 or S_2 is populated (Fig.10b Inset). Therefore, force clamp experiments within these force ranges show dissociation of slip type (Fig.3b dashed plots), thus yielding the corresponding dissociation rate constants (k_1, k_2) and bond lengths (x_1, x_2). In contrast, when performing force clamp experiments at intermediate forces, the HD/HS complex can dissociate either from S_1 **or** S_2. Hence, force clamp datasets obtained within this force regime comprise individual complex life times of both states and the average life time τ is therefore a superposition of S_1 and S_2.

Despite the small number of molecular systems exposing catch bond behavior there are several models describing the same. So far, there is a vital discussion about the interpretation of catch bond characteristics. Nevertheless, this exotic and unintuitive binding surely indicates a non-trivial energy landscape.

3 Summary and Outlook

The investigation of interactions between single molecules and the manipulation of structures on the molecular scale is an interdisciplinary task. Although commercial atomic force microscopes are readily available, exploiting the entire potential of these instruments requires detailed knowledge and experience. On the other hand, a general understanding of (bio-) chemical issues, the synthesis of tailor-made molecules and the preparation of specifically functionalized surfaces are just as important for successful experiments. But as diverse the requirements are, as multifaceted are the topics that can be addressed with this technique. The presented examples show that the binding in complexes with a wide range of affinities can be studied and dynamic processes as optomechanical switching can be followed.

There are a multitude of research areas that might get essential progress by the information accessible from single-molecule force spectroscopy experiments. Fundamental questions of theoretical physics might be addressed in studies on sophisticated biomolecular complexes and supramolecular structures. SMFS experiments on simple model complexes can help to develop mathematical models for macroscopic processes as adhesion, mechanical wear and (biological) recognition events. The impact of forces on interactions between the building blocks of (large) non-covalently linked aggregates, dynamic systems, chemical reactions and connections between surfaces mediated by supramolecular interactions is already under investigation. In the general view, single molecule force spectroscopy connects characteristic binding parameters of the macroscopic with the microscopic, molecular world.

References

[1] Fischer, E. *Berichte der deuschen chemischen Gesellschaft* (1894).
[2] Rief, M., Gautel, M., Oesterhelt, F.and Fernandez, J., and Gaub, H. *Science* **276**, 1109–1112 (1997).
[3] Dammer, U., Popescu, O., Wagner, P., Anselmetti, D., Güntherodt, H. J., and Misevic, G. N. *Science* **267**(5201), 1173–1175 (1995).
[4] Florin, E. L., Moy, V. T., and Gaub, H. E. *Science* **264**(5157), 415–417 (1994).
[5] Lee, G. U., Chrisey, L. A., and Colton, R. J. *Science* **266**(5186), 771–773 (1994).
[6] Evans, E. and Ritchie, K. *Biophys. J.* **72**(4), 1541–1555 Apr (1997).
[7] Bell, G. I. *Science* **200**(4342), 618–627 May (1978).
[8] Kramers, H. A. *Physica* **7**, 284–304 (1940).
[9] Smoluchowski, M. *Sitzungsberichte d. Kais. Ak. d. Wiss. in Wien. Math.-naturw. Klasse* (1915).
[10] Ritzefeld, M., Walhorn, V., Kleineberg, C., Bieker, A., Kock, K., Herrmann, C., Anselmetti, D., and Sewald, N. *Biochemistry* **52**(46), 8177–8186 Nov (2013).
[11] Wollschläger, K., Gaus, K., Körnig, A., Eckel, R., Wilking, S.-D., McIntosh, M., Majer, Z., Becker, A., Ros, R., Anselmetti, D., and Sewald, N. *Small* **5**(4), 484–495 Apr (2009).
[12] Lehn, J. M. *Science* **260**(5115), 1762–1763 Jun (1993).
[13] Lehn, J.-M. *Angewandte Chemie International Edition in English* **27**(1), 89–112 (1988).
[14] Nguyen, S. T., Gin, D. L., Hupp, J. T., and Zhang, X. *Proc Natl Acad Sci U S A* **98**(21), 11849–11850 Oct (2001).
[15] Anselmetti, D., Bartels, F. W., B., A., Decker, B., Eckel, R., McIntosh, M., Mattay, J., Plattner, P., Ros, R., Schafer, C., and Sewald, N. *Langmuir* **24**(4), 1365–1370 (2008). PMID: 18062710.
[16] Eckel, R., Ros, R., Decker, B., Mattay, J., and Anselmetti, D. *Angew. Chem. Int. Ed.* **44**(3), 484–488 (2005).
[17] Schäfer, C., Decker, B., Letzel, M., Novara, F., Eckel, R., Ros, R., Anselmetti, D., and Mattay, J. *Pure Appl. Chem.* **78**, 2247–2259 (2006).
[18] Schäfer, C., Eckel, R., Ros, R., Mattay, J., and Anselmetti, D. *J Am Chem Soc* **129**(6), 1488–1489 (2007).
[19] Dembo, M., Torney, D. C., Saxman, K., and Hammer, D. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **234**(1274), 55–83 Jun (1988).
[20] Marshall, B. T., Long, M., Piper, J. W., Yago, T., McEver, R. P., and Zhu, C. *Nature* **423**(6936), 190–193 May (2003).
[21] Bartels, F. W., McIntosh, M., Fuhrmann, A., Metzendorf, C., Plattner, P., Sewald, N., Anselmetti, D., Ros, R., and Becker, A. *Biophys J* **92**(12), 4391–4400 Jun (2007).
[22] Dammer, U., Hegner, M., Anselmetti, D., Wagner, P., Dreier, M., Huber, W., and Güntherodt, H. J. *Biophys. J.* **70**(5), 2437–2441 May (1996).
[23] Eckel, R., Wilking, S. D., Becker, A., Sewald, N., Ros, R., and Anselmetti, D. *Angew. Chem. Int. Ed.* **44**(25), 3921–3924 (2005).
[24] Fuhrmann, A., Schoening, J. C., Anselmetti, D., Staiger, D., and Ros, R. *Biophys J* **96**(12), 5030–5039 Jun (2009).
[25] Harder, A., Walhorn, V., Dierks, T., Fernández-Busquets, X., and Anselmetti, D. *Biophys J* **99**(10), 3498–3504 Nov (2010).
[26] Hinterdorfer, P., Baumgartner, W., Gruber, H. J., Schilcher, K., and Schindler, H. *Proc. Nat. Acad. Sci. U S A* **93**(8), 3477–3481 Apr (1996).
[27] Moy, V. T., Florin, E. L., and Gaub, H. E. *Science* **266**(5183), 257–259 (1994).
[28] Schröder, T., Geisler, T., Walhorn, V., Schnatwinkel, B., Anselmetti, D., and Mattay, J. *Phys Chem Chem Phys* **12**(36), 10981–10987 Sep (2010).
[29] Walhorn, V., Schäfer, C., Schröder, T., Mattay, J., and Anselmetti, D. *Nanoscale* **3**(11), 4859–4865 Nov (2011).
[30] Guo, B. and Guilford, W. H. *Proceedings of the National Academy of Sciences* **103**(26), 9844–9849 (2006).

[31] Harder, A., Möller, A.-K., Milz, F., Neuhaus, P., Walhorn, V., Dierks, T., and Anselmetti, D. *submitted* (2015).
[32] Kong, F., García, A. J., Mould, A. P., Humphries, M. J., and Zhu, C. *J Cell Biol* **185**(7), 1275–1284 Jun (2009).
[33] Thomas, W. E., Trintchina, E., Forero, M., Vogel, V., and Sokurenko, E. V. *Cell* **109**(7), 913 – 923 (2002).
[34] Yago, T., Lou, J., Wu, T., Yang, J., Miner, J. J., Coburn, L., López, J. A., Cruz, M. A., Dong, J.-F., McIntire, L. V., McEver, R. P., and Zhu, C. *J Clin Invest* **118**(9), 3195–3207 Sep (2008).
[35] Bartolo, D., Derényi, I., and Ajdari, A. *Phys Rev E Stat Nonlin Soft Matter Phys* **65**(5 Pt 1), 051910 May (2002).
[36] Evans, E., Leung, A., Heinrich, V., and Zhu, C. *Proc Natl Acad Sci U S A* **101**(31), 11281–11286 Aug (2004).
[37] Pereverzev, Y. V., Prezhdo, O. V., Forero, M., Sokurenko, E. V., and Thomas, W. E. *Biophys J* **89**(3), 1446–1454 Sep (2005).
[38] Cheung, L. S.-L., Raman, P. S., Balzer, E. M., Wirtz, D., and Konstantopoulos, K. *Phys Biol* **8**(1), 015013 Feb (2011).
[39] Pereverzev, Y. V. and Prezhdo, O. V. *Phys. Rev. E* **73**, 050902 May (2006).
[40] Pereverzev, Y. V., Prezhdo, O. V., and Sokurenko, E. V. *The Journal of Physical Chemistry B* **112**(36), 11440–11445 (2008). PMID: 18710275.
[41] Liu, F. and Ou-Yang, Z.-c. *Phys. Rev. E* **74**, 051904 Nov (2006).
[42] Barsegov, V. and Thirumalai, D. *Proc Natl Acad Sci U S A* **102**(6), 1835–1839 Feb (2005).
[43] Barsegov, V. and Thirumalai, D. *The Journal of Physical Chemistry B* **110**(51), 26403–26412 (2006). PMID: 17181300.
[44] Beste, M. T. and Hammer, D. A. *Proc Natl Acad Sci U S A* **105**(52), 20716–20721 Dec (2008).
[45] Pereverzev, Y. V., Prezhdo, E., and Sokurenko, E. V. *Biophys J* **101**(8), 2026–2036 Oct (2011).
[46] Suzuki, Y. and Dudko, O. K. *Phys. Rev. Lett.* **104**, 048101 Jan (2010).
[47] Thomas, W., Forero, M., Yakovenko, O., Nilsson, L., Vicini, P., Sokurenko, E., and Vogel, V. *Biophys J* **90**(3), 753–764 Feb (2006).
[48] Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., and Shaw, D. E. *Current Opinion in Structural Biology* **19**(2), 120 – 127 (2009). Theory and simulation / Macromolecular assemblages.
[49] Lee, T.-H., Gonzalez, J. I., and Dickson, R. M. *Proc. Natl. Acad. Sci.* **99**(16), 10272–10275 Aug (2002).
[50] Yang, W. Y. and Gruebele, M. *Nature* **423**(6936), 193–197 May (2003).

C 5 Solid-State NMR

O. Petrov and M. Vogel
Institut für Festkörperphysik
Technische Universität Darmstadt

Contents

1 Introduction 2

2 Nuclear Magnetic Resonance Phenomenon 3
2.1 Nuclear spins in a static magnetic field 3
2.2 Effects of an alternating magnetic field 4
2.3 NMR signal in time and frequency domains 5

3 Spin Interactions and NMR Spectra 7
3.1 Chemical shift 7
3.2 Quadrupolar interaction 10
3.3 Dipole-dipole coupling 12
3.4 Spin echoes 14

4 Effects of Molecular Dynamics in NMR 15
4.1 Line-shape analysis 15
4.2 Spin-lattice and spin-spin relaxation analyses 18
4.3 Two-dimensional NMR experiments 20
4.4 Magic-angle spinning 23

5 Further Reading 24

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

The nuclear magnetic resonance (NMR) phenomenon was independently discovered in 1945 by two groups of physicists around Purcell and Bloch, who received the Nobel Prize for this achievement. Since then, the method has developed into a very powerful characterisation tool. Substantial contributions to this success story were made by Ernst and Wüthrich, who became Nobel Laureates in 1991 and 2002, respectively. Today, the most prominent application of the phenomenon is probably magnetic resonance imaging as a non-invasive method for medical diagnosis, as honoured with a Nobel Prize for Lauterbur and Mansfield in 2003.

NMR experiments are based on interactions of spin-bearing nuclei with external magnetic fields, on the one hand, and internal magnetic or electric fields, on the other. The external magnetic fields are a static field $\mathbf{B}_0$ and an alternating field $\mathbf{B}_1$, which are often perpendicular to each other. The static one is usually applied by inserting the sample into a superconducting magnet, enabling typical field strengths of ca. 10 T. The $\mathbf{B}_0$ field serves to polarize the nuclear spin ensemble, resulting in an equilibrium magnetization $\mathbf{M}_0$. The alternating one is customarily generated by a coil wound around the sample. The $\mathbf{B}_1$ field is nowadays applied in form of short radio-frequency pulses, which allow one to manipulate the spin system and, hence, to create a non-equilibrium magnetization $\mathbf{M} \neq \mathbf{M}_0$. A typical NMR setup is depicted in Fig. 1.

By convention, $\mathbf{B}_0$ defines the z axis of the laboratory frame. While the external $\mathbf{B}_0$ and $\mathbf{B}_1$ fields are used to produce a desired NMR signal, the internal fields reflect the local properties of the sample. Hence, analyses of the strengths and fluctuations of the internal fields yield valuable information about the microscopic structure and dynamics of the studied sample. In the following, we will assume that the external fields are much stronger than the internal ones, which is true in the vast majority of cases.

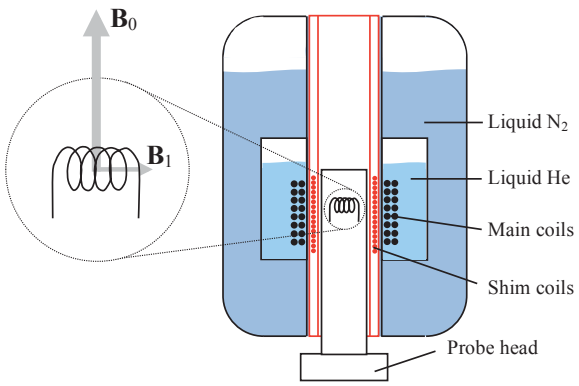


Fig. 1: *Sketch of an NMR setup: The sample is inserted into a coil, which in turn is placed into a superconducting magnet, producing a static magnetic field B_0 along the z axis of the laboratory frame. The sample coil, which is part of a resonant circuit, has a twofold task: It generates an alternating magnetic field B_1, applied in form of a short radio-frequency pulse, and it picks up the NMR signal in response to the associated excitation of the nuclear spin system. Shim coils are used to ensure a homogeneous field B_0 throughout the sample volume.*

Nuclei bearing a spin $\mathbf{I} \neq 0$, e.g., ^{1}H, ^{2}H, ^{7}Li, ^{13}C etc., possess a magnetic dipole moment,

$$\boldsymbol{\mu} = \gamma \mathbf{I} \tag{1}$$

with the nuclear-specific gyromagnetic ratio γ. In an external static magnetic field $\mathbf{B}_0$, nuclear magnetic moments tend to align, while thermal fluctuations disturb this alignment. Curie's law specifies the resulting nuclear magnetization $\mathbf{M}_0$ in thermal equilibrium (Section 2.1):

$$\mathbf{M}_0 = \frac{C}{T}\mathbf{B}_0 \tag{2}$$

Here, C denotes the Curie constant. An appropriate alternating magnetic field $\mathbf{B}_1$ deflects the nuclear magnetization from the equilibrium direction $\mathbf{M}_0 \parallel \mathbf{B}_0$ (Section 2.2), initiating a precession around $\mathbf{B}_0$. The angular frequency of this precession is called the Larmor frequency

$$\boldsymbol{\omega}_0 = -\gamma \mathbf{B}_0 \tag{3}$$

It takes on values in the radio-frequency range. The precessing magnetization $\mathbf{M}(t)$ induces a voltage in a coil around the sample, enabling a detection of an NMR signal (Section 2.3).

The time evolution $\mathbf{M}(t)$ is affected by various internal fields. As will be outlined in Section 3, the internal fields are produced by the local environments and, thus, they provide valuable insights into microscopic properties of a studied sample. In the course of time, thermal equilibrium is restored by relaxation processes within the nuclear spin system. While spin-spin relaxation causes a decay of the transversal magnetization components M_x and M_y, spin-lattice relaxation recovers the equilibrium value M_0 of the longitudinal magnetization component M_z. Fluctuations of the internal fields result from molecular dynamics. The effects of these fluctuations on spin-spin and spin-lattice relaxations as well as other NMR observables will be discussed in Section 4.

2 Nuclear Magnetic Resonance Phenomenon

2.1 Nuclear spins in a static magnetic field

The interaction between the nuclear magnetic moments and the static magnetic field is described by the Zeeman Hamiltonian

$$H_Z = -\boldsymbol{\mu}\mathbf{B}_0 = -\gamma I_z B_0 \tag{4}$$

Solving Schrödinger's equation for this Hamiltonian, we find that there are $2I+1$ states (called Zeeman eigenstates) characterized by well-defined spin z-angular momentum, where I is the spin quantum number. These Zeeman eigenstates are distinguished by the magnetic quantum number m, taking on values I, I–1, …, $-I$. The respective energies, or Zeeman eigenvalues, are

$$E_m = -\gamma \hbar B_0 m \tag{5}$$

Fig. 2 shows the Zeeman levels for I = ½ and I = 1. The energy gap between neighboring Zeeman eigenstates (Δm = 1) is proportional to the B_0 field strength:

$$\Delta E = \gamma \hbar B_0 \tag{6}$$

Thus, the Larmor frequency ω_0 and the energy difference ΔE are related according to

$$\omega_0 = \Delta E / \hbar \tag{7}$$

In thermal equilibrium, the population of the Zeeman eigenstates obeys a Boltzmann distribution. The equilibrium magnetization $\mathbf{M_0}$ results from the difference ΔN in the populations. For the example of ^{1}H with I = ½ and $\gamma > 0$, we obtain

$$\frac{N_{+1/2} - N_{-1/2}}{N_{+1/2}} = 1 - \exp(-\Delta E / k_B T) \tag{8}$$

As the energy difference ΔE is small compared to the thermal energy at ambient temperatures, there is only a slight difference in spin populations (~ 0.001% at 300 K). For $\Delta E << k_B T$, Curie's law is obtained from Eq. (8).

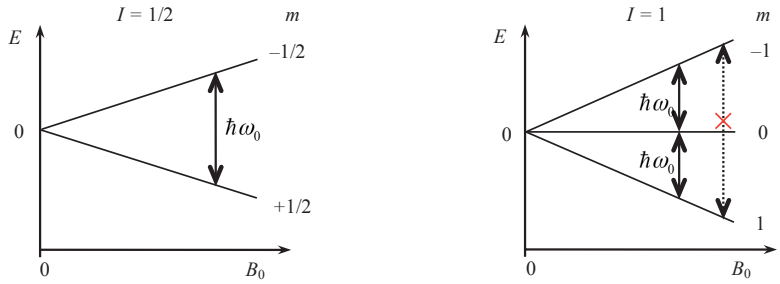


Fig. 2: *Zeeman energy levels for a system of spins I = ½ and I = 1, in the case of γ > 0. The solid arrows indicate transitions with Δm=±1 giving rise to an NMR signal. The dashed arrow marks forbidden transitions with Δm=±2.*

2.2 Effects of an alternating magnetic field

To induce spin transitions between the Zeeman eigenstates, the sample is irradiated by an alternating magnetic field $\mathbf{B}_1$ with a frequency matching the gap between the Zeeman energy levels, i.e., an alternating field that is resonant with the Larmor frequency, $\omega_{RF} \approx \omega_0$. It is usually applied in form of a short radio-frequency pulse using a coil, which is wrapped around the sample and part of a resonant circuit. Let us assume that the coil produces an alternating field $2B_1 \cdot \cos(\omega_{RF} t + \phi)$. This linearly polarized field can be regarded as a sum of two circularly polarized fields rotating in opposite directions, $B_1 \cdot \exp(\pm i\omega_{RF} t + \phi)$. The component, which is consistent with the spin precession, causes spin transitions, whereas the other component can be neglected. Hence, we are left with an effective rotating field $B_1 \cdot \exp(i\omega_{RF} t + \phi)$.

To understand the action of this rotating field, it is useful to transform to a coordinate system rotating with ω_{RF} around the z axis, where $\mathbf{B}_1$ appears static. The time evolution of the magnetization $\mathbf{M}$ in this rotating frame is determined by

$$\left(\frac{d\mathbf{M}}{dt}\right)_{rot} = \mathbf{M}\times\left[\left(\gamma B_0 - \omega_{RF}\right)\mathbf{e}_z + \gamma\mathbf{B}_1\right] = \mathbf{M}\times\gamma\mathbf{B}_{eff} \tag{9}$$

Evidently, in the rotating frame, $\mathbf{M}$ precesses around the direction of an effective field $\mathbf{B}_{\mathrm{eff}}$.

In resonance, $\omega_{RF} = \omega_0$, $\mathbf{B}_{\mathrm{eff}} = \mathbf{B}_1$, see Fig. 3 (left). Thus, a resonant radio-frequency pulse of duration Δ_p rotates the $\mathbf{M}$ vector from $\mathbf{B}_0 \parallel \mathbf{z}$ by the angle

$$\alpha = \gamma B_1 \Delta_p \tag{10}$$

In particular, resonant "90° pulses", i.e., pulses of length Δ_p corresponding to $\alpha = 90°$, flip the magnetization $\mathbf{M}$ from the z direction to the xy plane. The direction of $\mathbf{B}_1$ in the rotating frame is determined by its phase ϕ relative to this coordinate system. Thus, radio-frequency pulses of appropriate frequency, length and phase allow one to manipulate the macroscopic magnetization in a defined manner. Off resonance, $\omega_{RF} \neq \omega_0$, $\mathbf{B}_{eff}$ does not coincide with $\mathbf{B}_1$, see Fig. 3 (right). Thus, $\mathbf{M}$ turns around a tilted field $\mathbf{B}_{\mathrm{eff}}$. The more ω_{RF} deviates from ω_0 the weaker is the effect of the alternating magnetic field on the magnetization. As a consequence, the radio-frequency pulses exclusively affect spins with a specific value of γ, rendering NMR an isotope-selective method. For example, in ^{1}H and ^{2}H NMR, use of an appropriate frequency of the $\mathbf{B}_1$ field enables selective excitation of proton and deuteron spins, respectively.

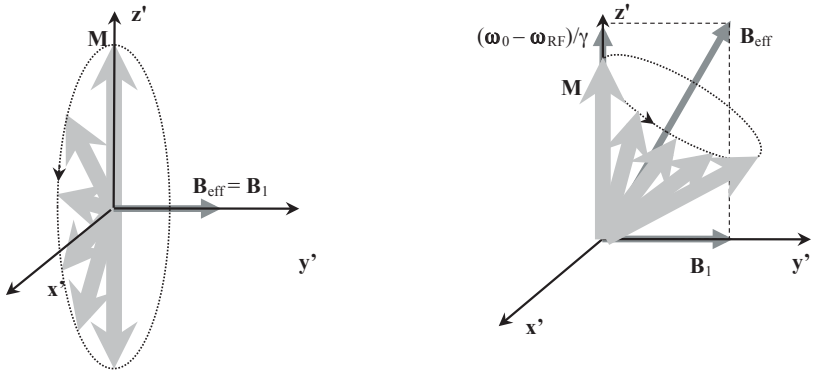


Fig. 3: *Precession of M during a radio frequency pulse, as seen from the coordinate system x'y'z' rotating at the frequency ω_{RF} with B_1 around B_0. Left: $\omega_{RF} = \omega_0$; the effective field in the x'y'z' frame is B_1. Right: $\omega_{RF} < \omega_0$; the effective field is $B_1+(\omega_0 - \omega_{RF})/\gamma$.*

2.3 NMR signal in time and frequency domains

After a radio-frequency pulse, a deflected magnetization $\mathbf{M}$ precesses around $\mathbf{B}_0$. This precession causes an induction signal in the sample coil, which is detected in the experiment. Usually, the same coil is utilized for pulse generation and signal detection. During precession individual spins gradually lose their phase coherence since local fields lead to somewhat different

precession frequencies. As a result, the transversal magnetization components, $M_x(t)$ and $M_y(t)$, decrease, a phenomenon called free induction decay (FID). Strictly speaking, the loss of a phase relation has irreversible and reversible parts, see Secs. 3.4 and 4.2.

For the present, we do not pursue this issue and assume that the decay of transversal magnetization is characterized by a single time constant T_2^*. Simultaneously, spin-lattice relaxation recovers the equilibrium value M_0 of the longitudinal magnetization component, M_z. In the simplest case, this relaxation is exponential with a characteristic time constant T_1. Then, the time evolution of **M** subsequent to a 90°$_y$ pulse, i.e., a 90° pulse along the y axis of the rotating frame (Fig. 3 (left)), is given by

$$\begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} = \begin{bmatrix} M_0 \sin\alpha \cos(\omega_0 t)\, e^{-t/T_2^*} \\ M_0 \sin\alpha \sin(\omega_0 t)\, e^{-t/T_2^*} \\ M_0 \cos\alpha + (1 - e^{-t/T_1})(M_0 - M_0 \cos\alpha) \end{bmatrix} \xrightarrow{\alpha = 90^\circ} \begin{bmatrix} M_0 \cos(\omega_0 t)\, e^{-t/T_2^*} \\ M_0 \sin(\omega_0 t)\, e^{-t/T_2^*} \\ M_0 (1 - e^{-t/T_1}) \end{bmatrix} \tag{11}$$

Because it is only a small frequency range around the Larmor frequency that is of interest in NMR experiments, see below, the NMR signal is de-modulated and filtered down to the difference frequency, $\omega_0 - \omega_{RF}$. Using quadrature detection, it is possible to detect the two components M_x and M_y separately. They are combined in the computer to give a complex FID signal

$$s(t) = M_x + iM_y \tag{12}$$

Fourier transformation allows one to convert the complex time signal $s(t)$ to the frequency domain, resulting in an absorption line and a dispersion line. For better spectral resolution, NMR spectra are presented as absorption lines. The time signal in Eq. (9) corresponds to a Lorentzian shape of the NMR spectrum, see Fig. 4.

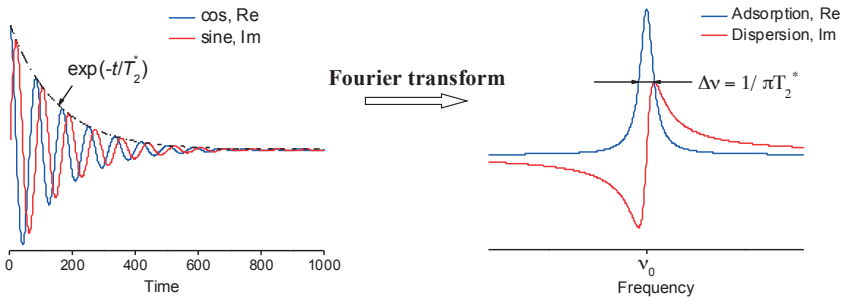


Fig. 4: *Complex FID signal with a frequency $\nu_0 = \omega_0/2\pi$ and its Fourier transform. Note the reciprocal relationship between the absorption line width (the spectrum width) $\Delta\nu$ and the signal decay time T_2^*.*

3 Spin Interactions and NMR Spectra

The nuclei in a sample do not only interact with the external magnetic fields $\mathbf{B}_0$ and $\mathbf{B}_1$, but, as aforementioned, also with internal magnetic and electric fields, which are determined by the properties of the local environments and contain valuable information about the microscopic structure and dynamics of the studied material. Since the external fields are much stronger than the internal fields in most cases, the latter cause relatively mild, but still resolvable shifts of the energy levels and resonance frequencies. As a result of the different local environments in a sample, there is usually not a single, but many resonance frequencies.

The power of NMR methods as analytic tools relies on the possibility to resolve effects resulting from the distributed and, in many cases, fluctuating internal fields. In general, the interactions of the nuclei with their local environments are anisotropic in nature and, hence, described by second rank tensors. While molecular reorientation is fast and isotropic in liquids so that it averages out the anisotropy of the interactions, such molecular dynamics is often absent in solids. Therefore, it is mandatory to consider the tensorial character of the spin interactions in NMR studies of solid-state samples. In such approaches, the orientation of the interaction tensor with respect to the laboratory frame determines the nuclear resonance frequency.

For the important case of molecular systems, the interaction tensor is linked to the molecular frame so that the NMR frequencies provide access to the molecular orientation relative to the $\mathbf{B}_0$ field. In the following, we will discuss three important examples of internal interactions: the shielding of the $\mathbf{B}_0$ field by electron clouds, the coupling of nuclear magnetic dipole moments, and the interaction of nuclear electric quadrupole moments with electric field gradients. We will not consider the *J* coupling, an electron-mediated spin interaction since it is a weak interaction (Fig. 5), which is of importance in liquid-state NMR, where motional averaging of internal interactions is effective, but plays no role in solid-state NMR.

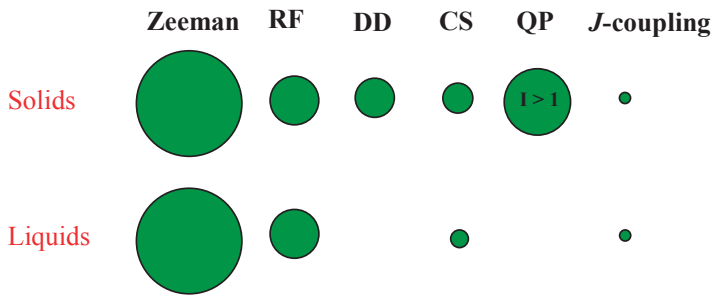


Fig. 5: *The rough relative strength of various spin interactions in NMR experiments.*

3.1 Chemical shift

The $\mathbf{B}_0$ field induces currents within electron orbitals around a nucleus. These diamagnetic currents generate a secondary field shielding the applied field in most cases (Fig. 6). Thus, the local field at the nucleus and the resonance frequency are reduced. This magnetic field shielding effect depends on the electron distribution around the nucleus and, hence, reflects the

chemical structure. Therefore, the related change of the resonance frequency is called chemical shift (CS). It scales with $\mathbf{B}_0$. The corresponding Hamiltonian reads:

$$H_{CS} = -\gamma \mathbf{I}\, \boldsymbol{\sigma}\, \mathbf{B}_0 \tag{13}$$

Here, $\boldsymbol{\sigma}$ is the CS tensor.

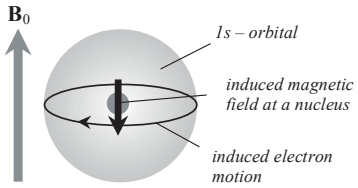


Fig. 6: *Diamagnetic currents in a 1s-orbital. In general, chemical bonding causes a non-spherical electron distribution around a nucleus, rendering the currents orientation dependent and, hence, resulting in chemical shift anisotropy.*

The corresponding small shift of the energy levels can be obtained from first-order perturbation theory. Hence, it is sufficient to consider parts that are diagonal in the basis of the Zeeman eigenfunctions. The resulting truncated CS Hamiltonian is given by:

$$H_{CS} = -\gamma I \sigma_{zz}^{LAS} B_0 \tag{14}$$

The superscript indicates that σ_{zz}^{LAS} is the *zz* component of the $\boldsymbol{\sigma}$ tensor in its laboratory axes system (LAS) representation. The associated frequency shift can be written as

$$\omega_{CS} = -\omega_0 \sigma_{zz}^{LAS} \tag{15}$$

Decomposing $\boldsymbol{\sigma}$ in a symmetric and an antisymmetric part, it can be shown that the former is relevant, while the latter can be disregarded because it does not produce a frequency shift. The symmetric part is most conveniently expressed in its principal axes system (PAS). When we exploit that bilinear forms are coordinate independent, we can use the identity

$$\sigma_{zz}^{LAS} = \mathbf{b}_0^{PAS} \boldsymbol{\sigma}^{PAS} \mathbf{b}_0^{PAS} \tag{16}$$

with $\mathbf{b}_0=\mathbf{B}_0/B_0$ and $\boldsymbol{\sigma}$ expressed in the PAS. Describing the orientation of $\mathbf{B}_0$ in the PAS by the polar coordinates (θ,ϕ) and denoting the principal values of $\boldsymbol{\sigma}$ as σ_{xx}^{PAS}, σ_{yy}^{PAS}, and σ_{zz}^{PAS}, we obtain from Eqs. (15) and (16):

$$\omega_{CS} = -\omega_0 \left[\sigma_{xx}^{\mathrm{PAS}} (\cos\phi \sin\theta)^2 + \sigma_{yy}^{\mathrm{PAS}} (\sin\phi \sin\theta)^2 + \sigma_{zz}^{\mathrm{PAS}} (\cos\theta)^2 \right] \tag{17}$$

Straightforward calculation shows that this can also be written as:

$$\omega_{CS} = -\omega_0 \left[\sigma_{iso} + \sigma_{aniso}(\theta,\phi) \right] \tag{18}$$

Here, the isotropic and anisotropic parts are given by:

$$\sigma_{iso} = \frac{1}{3}\left(\sigma_{xx}^{\mathrm{PAS}} + \sigma_{yy}^{\mathrm{PAS}} + \sigma_{zz}^{\mathrm{PAS}}\right) \tag{19}$$

$$\sigma_{aniso}(\theta,\phi) = \frac{\sigma_{zz}^{PAS} - \sigma_{iso}}{2}\left[3\cos^2\theta - 1 - \frac{\sigma_{yy}^{PAS} - \sigma_{xx}^{PAS}}{\sigma_{zz}^{PAS} - \sigma_{iso}}\sin^2\theta\cos(2\phi)\right] \tag{20}$$

where we used the convention $\left|\sigma_{yy}^{PAS} - \sigma_{iso}\right| \le \left|\sigma_{xx}^{PAS} - \sigma_{iso}\right| \le \left|\sigma_{zz}^{PAS} - \sigma_{iso}\right|$. As a consequence of the CS anisotropy, the resonance frequency depends, in general, on the orientation of the **σ** tensor with respect to the $\mathbf{B}_0$ field. Analogous orientation dependencies are obtained for the dipolar and quadrupolar interactions, which will be discussed in the next sections. Generally, these anisotropic contributions to the resonance frequencies can be written as

$$\omega_{aniso}(\theta,\phi) = \frac{\delta}{2}\left[3\cos^2\theta - 1 - \eta\sin^2\theta\cos(2\phi)\right] \tag{21}$$

For the CS interaction, the anisotropy parameter δ and the asymmetry parameter η amount to:

$$\delta = -\omega_0(\sigma_{ZZ}^{PAS} - \sigma_{iso}), \qquad \eta = \frac{\sigma_{yy}^{PAS} - \sigma_{XX}^{PAS}}{\sigma_{ZZ}^{PAS} - \sigma_{iso}} \tag{22}$$

Considering that the chemical shift is proportional to the $\mathbf{B}_0$ field strength, it is measured in relative units, parts per million (ppm), to enable a comparison of spectra from different magnets.

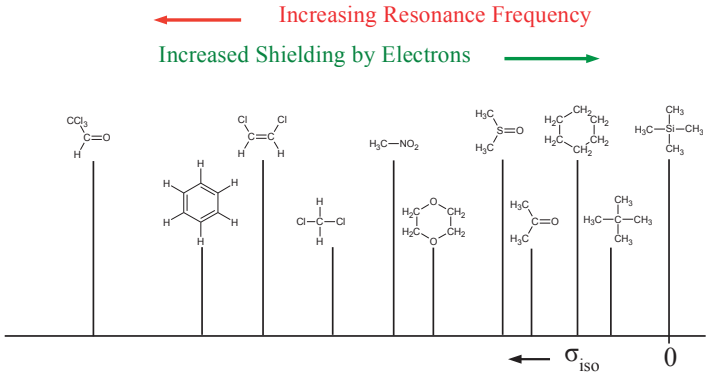


Fig. 7: *^{1}H chemical shifts in different compounds relative to the value of the standard reference tetramethylsilane. The value of σ_{iso} provides detailed insights into the chemical environment of a nucleus.*

Specifically, the chemical shift is usually given as a normalized difference between the observed resonance frequency and that of a reference compound, $(\omega - \omega_{ref})/\omega_{ref}$. For historical reasons, NMR spectra governed by the CS interaction are usually plotted with the CS increasing from right to left.

In liquids, fast and isotropic molecular dynamics averages the chemical shift anisotropy to zero so that the isotropic chemical shift is observed in the NMR spectra. The value of σ_{iso} is characteristic for the chemical environment of a nucleus, enabling a discrimination of different structural entities, see Fig. 7. This effect forms the basis for numerous applications in structure analysis. In solids, such motional averaging is often absent so that σ_{aniso} governs the NMR spectra. Then, the observed frequency yields information about the orientation of the molecular or crystal frames relative to the laboratory frame. For example, it may characterize the orientation of a chemical bond. NMR spectra resulting from such anisotropic NMR interactions will be discussed for the dipolar and quadrupolar interactions in the next sections.

3.2 Quadrupolar interaction

Nuclei with spin $I > 1/2$ (e.g. ^{2}H, ^{7}Li, ^{14}N, in all about 74%) exhibit a non-spherical distribution of the electric charge, giving rise to a nuclear electric quadrupole moment Q. The quadrupole moment interacts with an electric field gradient at the nuclear site, which is determined by the electronic environment of the nucleus. This quadrupolar (QP) interaction is described by the Hamiltonian

$$H_{QP} = \frac{eQ}{2I(2I-1)} \mathbf{I}\,\mathbf{V}\,\mathbf{I}, \qquad V_{\alpha\beta} = \left(\frac{\partial^2 \Phi}{\partial r_\alpha \partial r_\beta} \right) \tag{23}$$

Here, e is the elementary charge and the electric field gradient (EFG) tensor $\mathbf{V}$ is given by the second spatial derivatives of the electrostatic potential Φ.

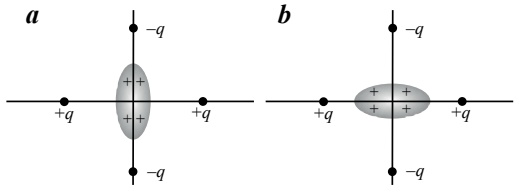


Fig. 8: *A quadrupole nucleus in an electric field gradient imposed by four charges. The quadrupole energy in case (a) is less than that in case (b).*

The shift of the Zeeman levels due to the QP interaction can be calculated in analogy to the treatment of the CS interaction in Section 3.1. A difference between these interactions results from the fact that the EFG tensor, contrary to the CS tensor, is traceless. As a consequence, the QP interaction, unlike the CS interaction, has no isotropic part. Thus, the QP interaction can only be observed in solids and the QP frequency amounts to $\omega_{QP} = \omega_{aniso}(\theta,\phi)$, as defined in Eq. (21), with anisotropy and asymmetry parameters:

$$\delta = k(I)\frac{eQeq}{\hbar} = k(I)C_Q, \qquad \eta = \frac{V_{xx}^{PAS} - V_{yy}^{PAS}}{eq} \tag{24}$$

Here, the custom convention $V_{zz}^{PAS} = eq$ is used. The constant *k(I)* indicates that the anisotropy parameter depends on the value of *I*, as one may expect based on the definition of H_{QP} in Eq. (23). For example, one has $k = 3/4$ for $I = 1$ (^{2}H, ^{6}Li etc.) and $k = 1/2$ for $I = 3/2$ (^{7}Li, ^{11}B etc.). Often, the quadrupolar coupling constant C_Q, rather than the anisotropy parameter δ is employed to characterize the nuclear-specific strength of the QP interaction.

The effect of the QP interaction on the Zeeman levels is shown for $I = 1$ and $I = 3/2$ in Fig. 9. Evidently, the QP interaction shifts the levels with $+m$ and $-m$ by the same amount. Consequently for $I = 3/2$, it does not affect the central transition $m = +1/2 \leftrightarrow m = -1/2$, but only the satellite transitions $m = \pm 3/2 \leftrightarrow m = \pm 1/2$. In general, there are $2I$ allowed ($\Delta m = \pm 1$) NMR transitions between the Zeemal levels.

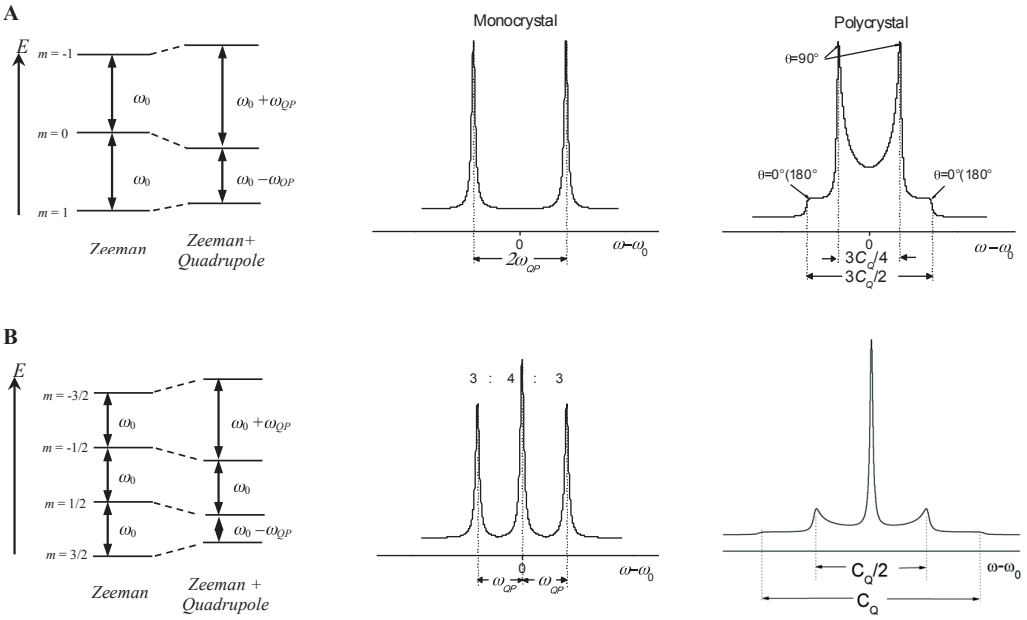


Fig. 9: *Energy levels for (A) I = 1 and (B) I = 3/2 together with the corresponding mono- and polycrystal spectra. For I = 3/2, the central transition is, in first order, not affected by the quadrupolar interaction so that the central peak is sharp even in a powder average.*

Let us consider the NMR spectra resulting from these energy diagrams for solid-state samples. For the moment, we assume that the observed nuclei occupy a single position in a crystal lattice. First, we discuss the situation for a moncrystal. In the case $I = 1$, we obtain two resonance lines, which are shifted by $+\omega_{QP}(\theta,\phi)$ and $-\omega_{QP}(\theta,\phi)$ relative to ω_0. Thus, the line positions and the line splitting are determined by the orientation of the EFG and, hence, of the crystal axes system relative to the laboratory frame. This effect can be exploited to map out the EFG by systematically rotating the crystal. In the case $I = 3/2$, a third line at ω_0 results from the central transition, which is, in first-order perturbation theory, not affected by the QP interaction. Now, we move on to a polycrystal. In this case, an isotropic distribution of crystal and, hence, tensor orientations exists so that the NMR spectrum reflects a powder average. For simplicity, we assume a symmetric EFG tensor, i.e., $\eta = 0$, leading to

$$\omega_{aniso}(\theta) \equiv \omega_{QP}(\theta) = \frac{\delta}{2}\left(3\cos^2\theta - 1\right) \tag{25}$$

so that the QP frequency solely depends on the angle θ between the z principal axis and the $\mathbf{B}_0$ magnetic field. Clearly, for an isotropic distribution, there are less principal axis parallel to the field direction ($\theta = 0°$) than those lying in the equatorial plane ($\theta = 90°$), as described by the size of the surface element, $\sin\theta$. Together with Eq. (25), this difference means that $\omega_{QP} = \delta$ is less often found in the NMR spectrum than $\omega_{QP} = -\delta/2$. Straightforward calculation yields the detailed line shape of the spectrum for a polycrystal. Fig. 9 shows the broad continuous spectra resulting for the examples of $I = 1$ and $I = 3/2$. Please note that frequency changes by $+\omega_{QP}$ and $-\omega_{QP}$ are obtained from different transitions between the Zeeman levels, leading to a symmetric line shape, e.g., to 'horns' at ω-$\omega_0 = \pm\delta/2$. For $I = 3/2$, the spectral contributions from the satellite transitions are broadened by the powder average, while that from the central transition is not, as the latter transition, unlike the former, is not affected by first-order QP interaction.

3.3 Dipole-dipole coupling

Another important internal spin interaction is the coupling of the related nuclear magnetic dipole moments (Fig. 10). For a pair of spins (j and k), this dipole-dipole (DD) coupling is described by the Hamiltonian:

$$H_{DD} = -\frac{\mu_0}{4\pi}\frac{\gamma_j\gamma_k}{r_{jk}^3}\hbar^2\left[3(\mathbf{I}^j\mathbf{e}_{jk})(\mathbf{I}^k\mathbf{e}_{jk}) - \mathbf{I}^j\mathbf{I}^k\right] \tag{26}$$

Here, the vector connecting the nuclei j and k is denoted as $\mathbf{r}_{jk}$ and the associated unit vector is $\mathbf{e}_{jk} = \mathbf{r}_{jk}/r_{jk}$. Thus, the DD coupling depends on the distance between the spins ($\sim r_{jk}^{-3}$) and on the orientation of the internuclear vector, as depicted in Fig. 11.

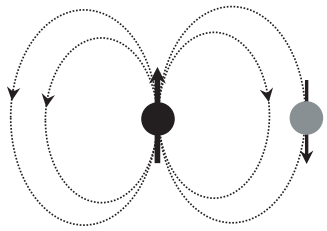


Fig. 10: *One nuclear dipole in a magnetic field by the other.*

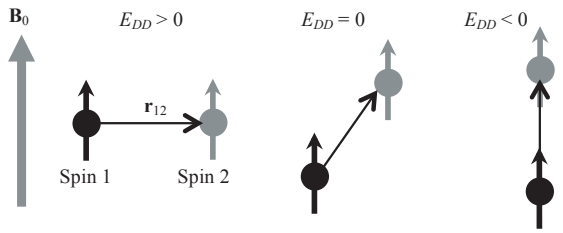


Fig. 11: *Orientation dependence of the dipole-dipole interaction energy E_{DD} in the case of a strong B_0 field.*

The effect of the DD interaction on the NMR spectrum can again be calculated using the treatment outlined in Section 3.1. It is found that, like the QP interaction, the DD interaction has no isotropic part and that, for an isolated spin pair, the DD coupling tensor is symmetric, i.e., $\eta = 0$. Thus, the orientation dependence of the frequency shift $\omega_{DD}(\theta)$ is given in analogy with Eq. (25). The truncation of the Hamiltonian in this first-order treatment leads, however,

to differences between homonuclear (e.g. ^{1}H-^{1}H) and heteronuclear (e.g. ^{1}H-^{13}C) spin couplings. Specifically, different spin parts of the resulting truncated Hamiltonians result in different anisotropy parameters for homonulcear and heteronuclear pairs:

$$\delta_{\text{hom}} = \frac{3}{2}\frac{\mu_0}{4\pi}\frac{\gamma_j^2}{r_{jk}^3}\hbar = \frac{3}{2}C_{DD}, \qquad \delta_{\text{het}} = \frac{\mu_0}{4\pi}\frac{\gamma_j\gamma_k}{r_{jk}^3}\hbar = C_{DD} \tag{27}$$

For example, the pair of protons in a water molecule has a distance of ~1.2 Å, resulting in $C_{DD} \approx 2\pi$ x 70 kHz. Due to the strong distance dependence, a determination of the DD coupling constant C_{DD} allows one to ascertain the spatial arrangement of spins, which, in turn, provides valuable insights into the structures of complex molecules, including proteins, or disordered solids, e.g., glasses, for which scattering techniques yield only limited information.

For a sample containing many spins, there is a DD coupling between each spin pair and, hence, the total Hamiltonian is obtained from summation over all these pairs. Thus, contrary to the CS and QP interactions, reflecting interactions of individual spins with their electronic environment, the DD coupling is a multi-spin interaction, which usually leads to broad unstructured NMR spectra. However, in view of the strong distance dependence of the DD interaction, a consideration of isolated spin pairs can still be a useful approximation in some situations, e.g., in the important case of diluted water molecules.

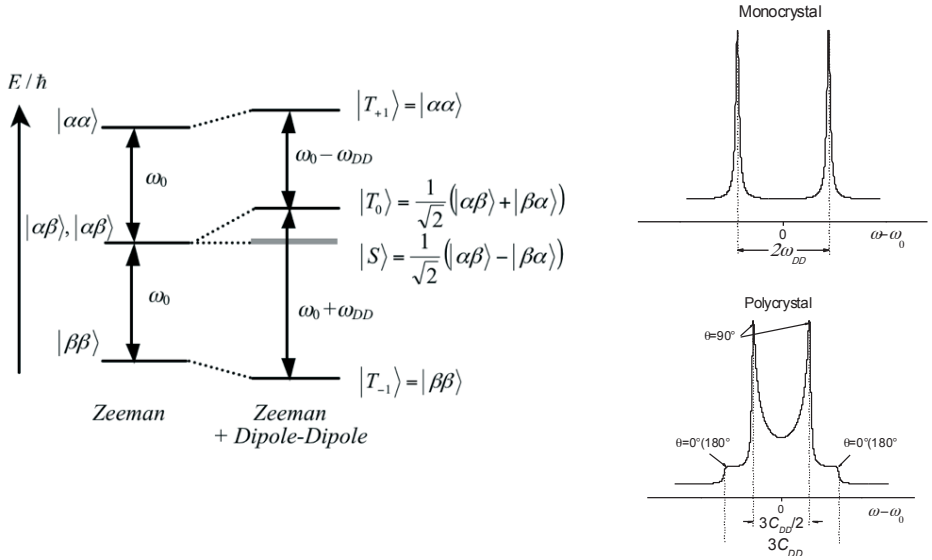


Fig. 12: *Energy levels of homonuclear spin pairs (I = ½) and their dipolar spectra. Note the entire resemblance of these spectra to the quadrupolar spectra of spins I=1 in Fig. 9.*

The different effect of homonuclear and heteronuclear DD couplings can be rationalized by the following arguments. First, we consider homonuclear spin pairs with $I = ½$, e.g., protons. A spin pair is a single quantum system and, as such, it can be described by Zeeman product states, $|\alpha\alpha\rangle$, $|\alpha\beta\rangle$, $|\beta\alpha\rangle$ and $|\beta\beta\rangle$. However, these product states are not the eigenstates

when DD couplings are relevant. Rather, there is a new set of stationary states $|S\rangle$ and $|T_M\rangle$ (Fig. 12). The states $|T_M\rangle$ correspond to the total angular momentum $I = 1$, with respective $M = 1, 0, -1$. They are called triplet states of a spin pair. The state $|S\rangle$ corresponds to $I = 0$ and is referred to as a singlet state. Due to the conservation of the total angular momentum during radio-frequency pulses, NMR transitions are only observed between the triplet states, leading to a doublet for the case of a monocrystal, as shown in Fig. 12.

In the case of $\omega_{DD} << \omega_0$, heteronuclear spin pairs can be with a good accuracy represented via Zeeman product states as stationary states. For comparison, a ^{13}C-^{1}H spin pair with $r_{12} = 0.11$ nm has $C_D = 2\pi\times(23$ kHz), whereas the difference in the Larmor frequencies of ^{13}C and ^{1}H in the typical magnetic field of 7 T is $2\pi\times(225$ MHz).

3.4 Spin echoes

In modern NMR experiments, it is extensively exploited that tailored sequences of radio-frequency pulses allow one to remove the effect of one interaction of the spin systems, while retaining that of another interaction. In general, it is necessary for a theoretical treatment to calculate the time evolution of the density matrix of the spin system under the action of the internal interactions and the applied pulses. Here, we restrict ourselves to a simple, but at the same time important example. We consider a system of isolated $I = 1/2$ spins subject to the CS interaction, which may differ for various spin subensembles. In this case, the time evolution can be illustrated in the framework of a vector model. In doing so, we exploit that the discussion is simplified when considering the situation in the rotating frame.

We study such spin system during the Hahn echo sequence (Fig. 13) featuring a 90° pulse, which, after a delay time t_1, is followed by a 180° pulse, i.e., by a twice as long pulse, leading to a rotation angle $\alpha = 180°$. For simplicity, we assume that both pulses have the same phase, i.e., the $\mathbf{B}_1$ fields point in the same (x) direction in the rotating frame. In the conventional short-hand notation, this pulse sequence reads $90°_x - t_1 - 180°_x - t_2$, where the numbers characterize the pulse lengths and the indices the pulse phases. In Fig. 13, the time evolution of the spin systems is depicted: Directly after the $90°_x$ pulse, the magnetization is aligned along the y axis of the rotating frame and all spins are in phase. During t_1, the spins precess with different frequencies as a consequence of diverse CS interactions, leading to a distribution of their magnetic moments over the xy plane. This de-phasing is accompanied a rapid decay of the NMR signal on a time scale T_2^*, see Section 2.3. The $180°_x$ pulse inverts the spins.

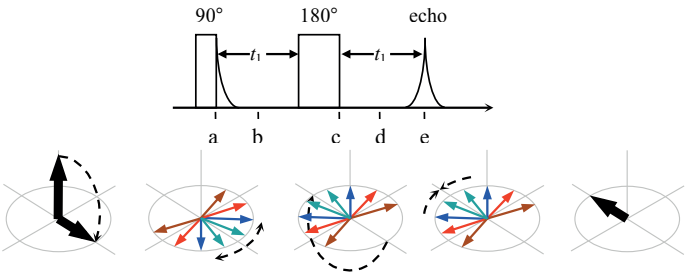


Fig. 13: *Formation of a Hahn echo at a time $t_2 = t_1$ after the second radio-frequency pulse.*

Hence, spins with higher frequencies find themselves 'behind' spins with lower frequencies, as the sense of the precession is conserved. During t_2, the former spins catch up with the latter ones. At $t_2 = t_1$, the re-phasing is complete and a spin echo forms, leading to a recovery of the NMR signal. This phenomenon is known as Hahn echo, named after its discoverer Erwin Hahn.

Such echo forms when the spins do not exchange their precession frequencies during the de- and rephrasing periods. The echo amplitude decreases when molecular dynamics causes random fluctuations of the precession frequencies. Thus, the reversible loss of transversal magnetization, which can be overcome in echo experiments, should not be mistaken with the irreversible loss, which results, in particular, from molecular motion and manifests itself in a decay of the echo amplitude, see Section 4.2. The latter phenomenon is known as spin-spin relaxation. In simple cases, it is described by an exponential with a characteristic time constant T_2.

Spin echoes are of great practical importance for the measurement of undistorted solid-state spectra. As the same coil is usually employed for pulse generation and signal detection, the weak NMR signal is not measurable during the decay of the strong $\mathbf{B}_1$ field after a radio-frequency pulse, leading to a 'dead time' of the receiver system, which typically lasts for a few microseconds. For solid samples, the NMR signals decay rapidly (T_2^*) so that substantial parts get lost in this dead time, leading to distorted line shapes. To overcome this problem, spin echoes are used to refocus the NMR signal outside the dead time. When molecular dynamics can be neglected, the time evolution after the echo maximum corresponds to that after the 90° pulse and, hence, Fourier transformation of the former NMR signal provides access to undistorted line shapes.

4 Effects of Molecular Dynamics in NMR

Various NMR methods provide insights into molecular motion. As the time windows of these techniques differ, a broad dynamic range can be covered in combined approaches. In the following, we discuss spin-lattice relaxation measurements, which are most sensitive to motions in the nanosecond regime. Moreover, we elucidate line-shape and stimulated-echo analyses providing access to dynamics in the microsecond and millisecond ranges, respectively.

4.1 Line-shape analysis

In general, the above discussed solid-state line shapes change when the observed nuclei exchange their resonance frequencies during signal detection. Such frequency exchange can be caused by molecular dynamics or chemical exchange. The relevant time scale is the time interval during which spin species with diverse frequencies dephase by ~1 rad and, hence, the FID disappears. For a spectrum spread over $\Delta\omega$, this time interval is $\sim 1/\Delta\omega$. In the case of polycrystalline or disordered samples, the width of the spectrum is determined by the anisotropy parameter, resulting in a characteristic time scale $\sim 1/\delta$ for line-shape measurements.

First, we consider that nuclei may exchange between chemically distinguishable sites so that they experience different internal fields, e.g., diverse chemical shifts, during FID acquisition. Such situation may be found during a chemical reaction. The effect of the chemical exchange

on the spectrum depends on the relation between the separation of resonance lines associated with the exchangeable sites, $\Delta\omega$, and the rate of the chemical exchange, k_{ex}. The exchange rate usually increases with increasing temperature, resulting in characteristic line-shape changes, which are shown in Fig. 14. At sufficiently low temperatures, when $k_{ex} << \Delta\omega$, the lines of distinguishable sites are well resolved. As k_{ex} becomes comparable to $\Delta\omega$ upon heating, the individual lines first broaden and then coalesce into a single peak at an intermediate position between the individual lines. At sufficiently high temperatures, when $k_{ex} >> \Delta\omega$, the average peak sharpens when the exchange rate further rises. Thus, from line-shape analysis, one can estimate chemical reaction rates at equilibrium conditions, which are difficult to access otherwise.

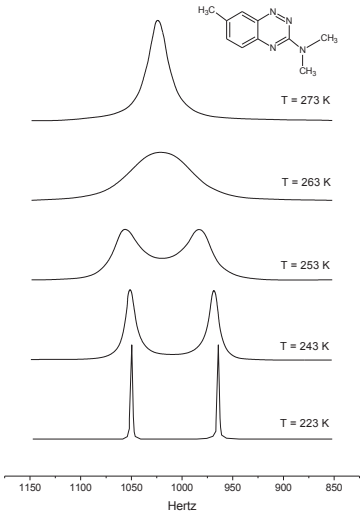


Fig. 14: : *^{1}H NMR spectra at 300 MHz of the N-methyl signals in a derivative of azapropazone as a function of temperature. The coalescence point (the spectrum at 263 K) is observed at $k_{ex} \approx \Delta\omega$ [Prog. NMR Spec. 43 (2003) 63–103].*

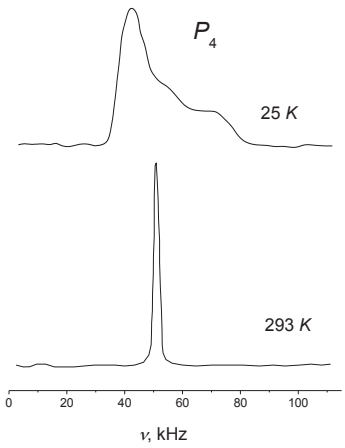


Fig. 15: : *^{31}P NMR spectra of solid white phosphorus at B_0 = 5.6 T. The spectrum at 25 K is dominated by the anisotropy of the CS interaction. Some blurring of the line shape results from the much weaker DD interactions. At room temperature, rapid reorientation of P_4 molecules averages out the anisotropy of the CS interaction [Chem. Phys. 6 (1974), 226-234].*

NMR spectra can also be affected by molecular dynamics. Then, the observed line shape depends on both the rate and the geometry of the motion. Isotropic motions characterized by short correlation times $\tau << 1/\delta \approx 1$ μs completely average out the anisotropy of the internal spin interactions. Such situation is found in liquids where the molecules show fast isotropic reorientation. Therefore, the NMR spectra of liquids exhibit sharp Lorentzian lines at frequencies corresponding to the centers of gravity of given spin interactions. As aforemen-

tioned, unlike the CS interaction, the DD and QP interactions do not have an isotropic part and, as such, they do not manifest themselves in NMR spectra of liquids.

Motional narrowing of NMR spectra is not restricted to liquids, but is can also be found in solids. Molecular dynamics in solids is, however, anisotropic rather than isotropic in most cases. Then, the line-shapes resulting from motional averaging yield valuable insights into the geometry of the dynamical process under study. As first example, Fig. 15 shows ^{31}P NMR spectra of solid white phosphorus at two temperatures. For this $I = ½$ spin (no QP interaction), a CS interaction stronger than DD couplings governs the experimental results. At a low temperature, a broad spectrum results from the anisotropy of the CS interaction because motion on the NMR time scale is absent, i.e., $\tau >> 1/\delta$. At a high temperature, a narrow line is observed as a consequence of dynamics with a correlation time $\tau << 1/\delta$. Specifically, P_4 molecules undergo fast rotational jumps between tetrahedral positions. This motion with cubic symmetry effectively averages out the anisotropy of the CS interaction, resulting in a liquid-like spectrum.

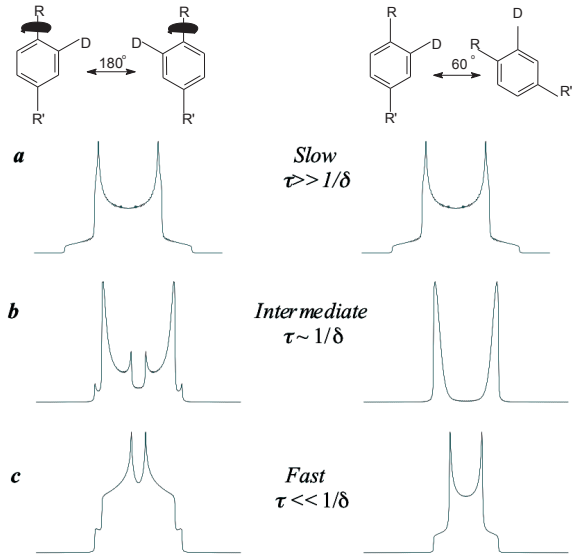


Fig. 16: *^{2}H NMR powder spectra of an aromatic molecule showing (left) 180° ring flips and (right) ring rotations in steps of 60°. The line shapes for different correlation times are compared.*

When fast ($\tau << 1/\delta$) molecular dynamics exhibits substantial anisotropy, as observed, e.g., in liquid crystals, the internal spin interactions retain an orientation dependence, but the tensor components of the static cases are replaced by tensor components, which are partially averaged by the motion. Hereby, the geometry of the dynamics determines the extent of the averaging. To illustrate the effect, we present ^{2}H NMR spectra ($I = 1$) of an aromatic molecule, which are dominated by the QP interaction, see Fig. 16. Specifically, we compare line shapes resulting from 180° flips of the ring and from ring rotations in 60° steps. We see that the motional-averaging effect is clearly different for these anisotropic motions. In the intermediate regime between slow and fast motion, $\tau \sim 1/\delta$, there is a particularly strong dependence on the

rate and geometry of the motion so that line-shape analysis yields a wealth of information about the dynamics.

4.2 Spin-lattice and spin-spin relaxation analyses

Valuable insights into molecular dynamics are also available from spin-lattice relaxation analysis. The disturbance of spin interactions by thermal motion induces spin transitions, which eventually bring spin-state populations back to thermal equilibrium. The transition probability and, hence, the spin-lattice relaxation rate $1/T_1$ depends on the squared amplitude and the spectral density of the motion-induced fluctuations of the relevant spin interaction. Hereby, the spectral density $J(\omega)$ is the Fourier transform of the correlation function $C(t)$ describing the fluctuations, see Fig. 17. Motions with a higher spectral density at the resonance frequency, $J(\omega_0)$, are more effective for spin-lattice relaxation. To illustrate the circumstances, consider a simple case of uncoupled $I = ½$ spins that move between sites with different local fields B_{loc}. The relaxation rate $1/T_1$ is then given by the product of the mean square fluctuating interaction $<(\gamma B_{loc})^2>$ and the spectral density at the resonance frequency according to:

$$1/T_1 = \gamma^2 \left\langle B_{loc}^2 \right\rangle J(\omega_0) \tag{28}$$

The form of $J(\omega)$ is determined by the motional process. For the simple case of isotropic diffusion, $C(t)$ is an exponentially decaying function,

$$C(t) = \mathrm{e}^{-t/\tau} \tag{29}$$

such that $J(\omega)$ takes a Lorentzian form

$$J(\omega) = \frac{\tau}{1 + \omega^2 \tau^2} \tag{30}$$

From Eq. (30), it can be seen that the spectral density at a given Larmor frequency ω_0 and, hence, the spin-lattice relaxation rate $1/T_1$ is a maximum when $\omega_0 \tau = 1$. This means that temperature-dependent relaxation times T_1 show a minimum when the correlation time τ of the observed molecular dynamics crosses the time scale $1/\omega_0$. Thus, for typical Larmor frequencies in high magnetic fields, a T_1 minimum indicates molecular dynamics in the nanosecond regime.

The exact mechanism of spin-lattice-relaxation is determined by the dominating spin interaction. Let us consider a molecular system. When CS or QP interactions dominate, spin-lattice relaxation probes the reorientation of single molecules. By contrast, when DD interactions are relevant, the relaxation is related to a random change of the distance vectors joining spin pairs. This gives rise to intra- and intermolecular inputs as well as rotational and translational contributions. Also, it should be stressed that the particular relation between $1/T_1$ and $J(\omega)$, see Eq. (23), depends on the type of the fluctuating interaction. For example, when DD or QP interactions dominate, the relaxation rate depends on not only $J(\omega_0)$, but also $J(2\omega_0)$. Furthermore, it should be mentioned that, in the case of complex molecular dynamics, the spectral density deviates from the Lorentzian form so that more elaborate functional forms are required for spin-lattice relaxation analysis.

Finally, we would like to point out that spin-lattice relaxation can be non-exponential for complex molecular dynamics or spin systems. One example is the case of complex dynamics with a distribution of correlation times τ, which leads to a distribution of relaxation times T_1, unless there is sufficiently fast exchange between spins relaxing at different rates.

Such exchange can, e.g., result from a fast transfer of magnetization via spin diffusion, i.e., via flip-flop processes of coupled spins. Another example is the case of a system of $I = 3/2$ spins, where QP fluctuations can induce satellite transitions, but not central transitions, see Fig. 9, so that different spectral components achieve equilibrium intensities at different rates.

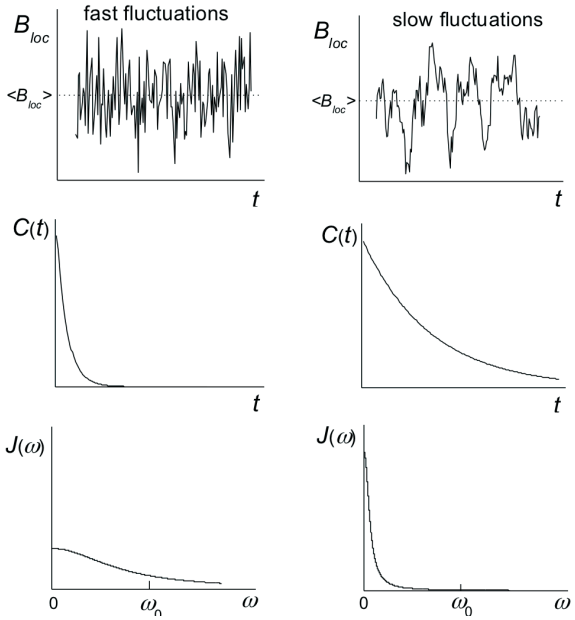


Fig. 17: *Correlation function C(t) and spectral density J(ω) for fast and slow fluctuations.*

To measure spin-lattice relaxation, the nuclear magnetization can be disturbed by inverting it with a 180° pulse (Fig. 18). The following return to thermal equilibrium can be monitored, when applying a 90° pulse after a variable relaxation delay t_d, which allows one to read out the instantaneous value of the z magnetization, $M_z(t_d)$, based on the amplitude of the resulting FID signal. As an alternative to this inversion-recovery experiment, the saturation-recovery experiment allows one to ascertain the spin-lattice relaxation of samples exhibiting $T_1 > T_2$, as found for solids.

In the latter experiment, the 180° pulse is replaced by a saturation sequence, i.e., by an appropriate pulse sequence that completely destroys the nuclear magnetization. In this way, it is possible to generate a well-defined non-equilibrium state of the magnetization independent of the history so that, unlike in the inversion-recovery experiment, it is not necessary to start from thermal equilibrium. The latter fact can be exploited to reduce the experimental time. In the simple case of exponential spin-lattice relaxation, the recovery curves obtained from these experiments are fitted with the functions:

$$M_z(t_d) = M_0(1 - 2e^{-t_d/T_1}) \quad \text{inversion-recovery} \tag{31}$$

$$M_z(t_d) = M_0(1 - e^{-t_d/T_1}) \quad \text{saturation-recovery} \tag{32}$$

yielding the spin-lattice relaxation time T_1.

Spin-spin relaxation describes the phenomenon that precessing spins lose their phase coherence as a consequence of randomly changing local fields. Unlike spin-lattice relaxation, spin-spin relaxation does not involve an energy transfer to other degrees of freedom, the 'lattice'. A spin dephasing results in a decay of the M_x and M_y components, rationalizing the synonym transversal relaxation. The characteristic time scale of spin-spin relaxation is called T_2.

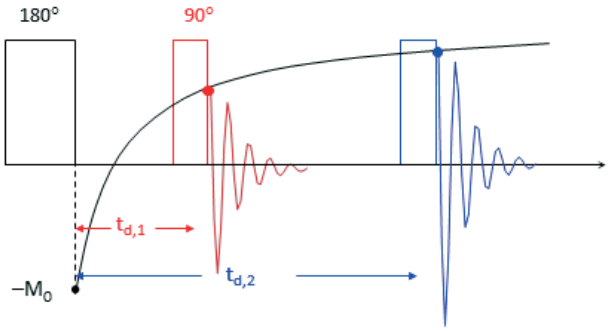


Fig. 18: *The inversion-recovery experiment to measure spin-lattice relaxation. To map out the recovery of the magnetization, the experiment is repeated for various delays t_d between the 180° and 90° pulses, which serve to invert the equilibrium magnetization and to read out the instantaneous magnetization, respectively.*

Similarly to the case of spin-lattice relaxation, the fluctuating spin interactions can be caused by molecular dynamics and, indeed, $T_2 \approx T_1$ in liquids. In solids, flip-flop transitions of dipolar coupled spins also modulate the local fields in a random fashion, leading to $T_2 << T_1$. Measurements of spin-spin relaxation usually involve an analysis of echo amplitudes. To follow the relaxation, it is possible to perform experiments with variable echo delays or with an increasing number of refocussing periods.

4.3 Two-dimensional NMR experiments

Two-dimensional (2D) NMR experiments allow one to ascertain slow frequency changes, providing access to slow molecular dynamics in solids. In these experiments, one correlates the resonance frequencies (ω_1 and ω_2) of a nucleus at two times. Such correlation can be achieved using a stimulated-echo pulse sequence, where two periods of frequency detection (t_1 and t_2), are separated by a mixing time $t_m >> t_1, t_2$, see Fig. 19.

In many cases, the resonance frequencies of the observed spins depend on the molecular orientations or positions. Then, the respective resonance frequencies of a nucleus during t_1 and t_2 are the same, $\omega_1 = \omega_2$, when molecular dynamics is absent, while $\omega_1 \neq \omega_2$ results from molecular rotational or translational motions with correlation times $t_1, t_2 << \tau \leq t_m$.

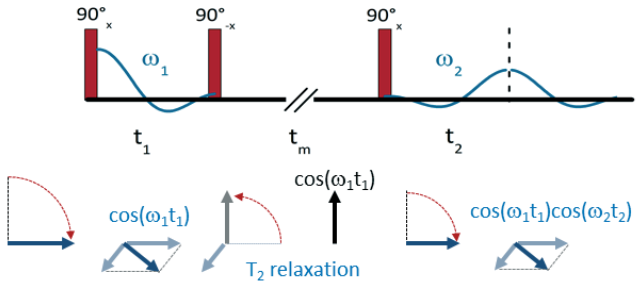


Fig. 19: *A three-pulse sequence with a stimulated echo at $t_2 = t_1$. In the lower picture, we visualize the time evolution for a particular subensemble of spins featuring a resonance frequency ω_1 during the evolution time t_1 and a resonance frequency ω_2 during the detection time t_2. Blue and black arrows mark transversal and longitudinal magnetization, respectively, faint arrows indicate the respective components of the magnetization, and red arrows illustrate the action of the pulses.*

To elucidate the working principle of stimulated-echo sequences in more detail, we consider the example of a molecular system featuring isolated $I = ½$ spins subject to the CS interaction. We suppose a single chemical environment of the observed spins, i.e., all interaction tensors are described by the same principal values and, hence, by the same anisotropy and asymmetry parameters, where we assume $\delta \neq 0$ and $\eta = 0$ for simplicity, while various molecular and, hence, tensor orientations exist.

In Fig. 19, we visualize the time evolution of such spin system during the stimulated-echo sequence, $90°_x - t_1 - 90°_{-x} - t_m - 90°_x - t_2$, considering a vector model in the rotating frame. The first pulse flips the magnetization from the z axis onto the y axis. During the subsequent evolution time t_1, the spins precess with various frequencies ω_1, which are determined by the respective molecular orientations. Hence, they accumulate different phases, leading to a modulation of the x and y components by $\sin(\omega_1 t_1)$ and $\cos(\omega_1 t_1)$, respectively. Depending on the phase of the second pulse, either of these components can be transformed into z magnetization. In our case, the second pulse ($90°_{-x}$) does not affect the x component, while it brings the modulated y component of the magnetization back onto the z axis. The remaining transversal x component decays by spin-spin relaxation, whereas the generated longitudinal z component does not so that it can be stored. The latter component is only affected by spin-lattice relaxation with $T_1 >> T_2$. After a mixing time $T_2 << t_m << T_1$, we are left with modulated z magnetization. The third pulse flips this magnetization onto the y axis. During the following detection time t_2, the spins again precess with orientation-dependent frequencies ω_2. Thus, when detecting the y component of the magnetization during this period, we measure a time signal

$$F_{cc}(t_1, t_2; t_m) \propto \langle \cos(\omega_1 t_1) \cos(\omega_2 t_2) \rangle \tag{33}$$

Appropriately adapting the pulse phases, it is also possible to record an analogous time signal

$$F_{ss}(t_1, t_2; t_m) \propto \langle \sin(\omega_1 t_1) \sin(\omega_2 t_2) \rangle \tag{34}$$

We see that the resonance frequencies and, thus, the molecular orientations at two times are correlated. Here, the pointed brackets denote the ensemble average. Moreover, the notation F_{ss}

(t_1,t_2; t_m) indicates that the signals $F_{\xi\xi}$ (t_1,t_2) parametrically depend on the value of t_m ($\xi\xi$ = *cc*, *ss*). We note that the pulse lengths and phases to be used in stimulated-echo sequences depend on the spin quantum number and the dominant spin interaction. For example, for I = 1 and QP interaction, measurements of F_{ss} involve a 90° pulse followed by two 45° pulses. For an understanding of this difference, it is necessary to calculate the time evolution of the density matrix of the spin system.

To justify the name stimulated-echo sequence, we assume $\omega_1 = \omega_2 \equiv \omega_{CS}$ and use trigonometric relations to rewrite Eq. (33) and Eq. (34) as

$$F_{\xi\xi}(t_1,t_2;t_m) \propto \left\langle \frac{1}{2}\cos\left[\omega_{CS}(t_2 - t_1)\right] \pm \frac{1}{2}\cos\left[\omega_{CS}(t_2 + t_1)\right] \right\rangle \tag{35}$$

Here, the plus and minus signs result for F_{cc} and F_{ss}, respectively. The value of the second term depends on ω_{CS} and, hence, there is destructive interference of the contributions from spins with different resonance frequencies. By contrast, the argument of the first term vanishes at $t_1 = t_2$, independent of the value of ω_{CS}, and, thus, there is constructive interference of identical contributions (½ cos(0) = ½) from different spins, manifesting itself in an echo. This phenomenon at $t_1 = t_2$ is known as stimulated echo.

F_{cc} and F_{ss} can be analyzed in the frequency or time domains. In both cases, straightforward analysis will be possible if molecular dynamics during t_1 and t_2 can be neglected. This condition, which will be assumed in the following discussion, is met for $1/\delta << \tau$, because the signals during these periods decay on the time scale of the inverse spectral width. Altogether, the typical time window of 2D NMR ranges from about 1-10 μs to 1-10 s, limited by $1/\delta$ and T_1, respectively.

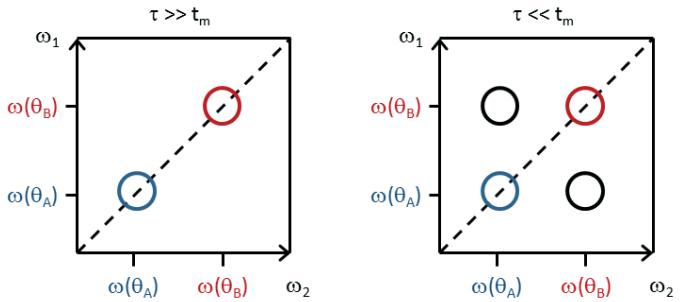


Fig. 20: *2D NMR spectra for rotational jumps between two possible orientations (θ_A and θ_B) of equal probability: (left) $\tau >> t_m$ and (right) $\tau << t_m$. In the former case, the spectral intensity is restricted to the diagonal. In the latter case, on-diagonal and off-diagonal peaks have the same intensity because there is a statistical redistribution of the molecular orientations after sufficiently long times.*

In frequency-domain analysis, an appropriate superposition of $F_{cc}(t_1,t_2; t_m)$ and $F_{ss}(t_1,t_2; t_m)$ is measured for various t_1 and t_2 values, but constant t_m, such that a 2D NMR absorption spectrum $S(\omega_1,\omega_2; t_m)$ can be obtained from Fourier transformation with respect to t_1 and t_2. The interpretation of this 2D spectrum is straightforward: $S(\omega_1,\omega_2; t_m)$ is proportional to the joint

probability density of finding a frequency ω_1 before the mixing time t_m and a frequency ω_2 after this period. For $\tau >> t_m$, the molecular orientations and, thus, the resonance frequencies do not change during the mixing time, leading to a 2D spectrum with intensity restricted to the diagonal $\omega_1 = \omega_2$. For $\tau \leq t_m$, molecular reorientation, in general, results in $\omega_1 \neq \omega_2$ and, hence, in a 2D spectrum with off-diagonal intensity, the shape of which reflects the motional geometry. In Fig. 20, the situation is illustrated for the simple case of rotational jumps between two possible molecular orientations of equal probability.

In time-domain analysis, one measures the amplitude of the stimulated echo at $t_1 = t_2 \equiv t_p$ as a function of the mixing time t_m. The resultant stimulated-echo decays map out two-time rotational correlation functions, $F_{ss}(t_m;t_p)$. These correlation functions parametrically depend on the length of the evolution time t_p, which determines the angular resolution of the experiment. In our case $\eta = 0$, where $\omega_{1,2} \propto P_2(\cos\theta)$, see Eq. (25), $F_{ss}(t_m;t_p)$ provides access to the correlation function of the second Legendre polynomial $P_2(\cos\theta)$ in the limit $t_p \rightarrow 0$:

$$F_{ss}(t_m;t_p \rightarrow 0) \approx t_p^2 \langle \omega_1 \omega_2 \rangle \propto \langle P_2[\cos\theta(t=0)] P_2[\cos\theta(t=t_m)] \rangle \tag{36}$$

4.4 Magic-angle spinning

Unlike in liquids, the anisotropy of spin interactions is not averaged out by fast and isotropic molecular reorientation in solids, resulting in broad line shapes. This broadening hampers, e.g., a discrimination of contributions from nuclei can be different chemical environments. To overcome this problem, the molecular reorientation is mimicked by rapid spinning of the sample. Specifically, the sample is rotated with frequencies in the kHz regime at the magic angle $\theta_s = 54.7°$ with respect to the direction of the $\mathbf{B}_0$ field. For this angle, $3\cos^2\theta - 1 = 0$ so that the anisotropy of the spin interactions is averaged out, see Eq. (25), provided the spinning frequency exceeds the spectral width. This technique is called magic angle spinning (MAS). It allows one to record liquid-like spectra for solids. In particular, narrow lines at positions given by the isotropic chemical shift can be observed. Therefore, MAS NMR is a versatile tool to ascertain structural properties of solids, e.g., of glasses, for which scattering techniques provide only limited insights.

5 Further Reading

General NMR textbooks:

1. A. Abragam, *"The principles of Nuclear Magnetism"*, Clarendon Press, Oxford, 1961
2. C. P. Slichter, *"Principles of Magnetic Resonance"*, Springer, Berlin, 1989
3. M. H. Levitt, *"Spin Dynamics"*, Wiley, Chichester, 2008

High-resolution NMR and spectral analysis:

4. H. Günther, *"NMR-Spektroskopie"*, Georg Thieme Verlag, Stuttgart, 1983
5. A. E. Derome, *"Modern NMR Techniques for Chemistry Research"*, Pergamon Press, Oxford, 1987

Solid-state NMR:

6. K. Schmidt-Rohr, H. W. Spiess, *"Multidimentional Solid-State NMR and Polymers"*, Academic Press, London, 1994

Spin relaxation:

7. R. Kimmich, *"NMR Tomography, Diffusometry, Relaxometry"*, Springer-Verlag, Berlin, Heidelberg, 1997
8. J. Kowalewski, L. Mäler, *"Nuclear Spin Relaxation in Liquids: Theory, Experiments and Applications"*, CRC Press, Boca Raton, Florida, 2006

On-line tutorials:

9. J. Keeler, *"Understanding NMR Spectroscopy"*, http://www.keeler.ch.cam.ac.uk/lectures/
10. J. P. Hornak, *"The basics of NMR"*, http://www.cis.rit.edu/htbooks/nmr/nmr-main.htm

C 6 Rheology of Complex Fluids

J. Vermant and Th. Schweizer
Department of Materials
ETH Zürich, Switzerland

Contents

1 Introduction **3**

2 Rheological Phenomena **5**
2.1 Viscoelasticity 5
2.2 Non-Newtonian behaviour 5
2.3 Yield stress and thixotropy 6
2.4 Elastic effects 7
2.5 Elongational viscosity 8

3 Linear viscoelasticity **8**
3.1 Creep 9
3.2 Stress Relaxation 10
3.3 Small Strain Oscillatory Experiments 11
3.4 Linear viscoelastic modelling 12

4 Non-Linear Viscoelasticity **17**
4.1 Non-linear relaxation modulus 17
4.2 Large amplitude oscillatory shear flow 18
4.3 Normal stress differences in shear flow 18
4.4 Shear thinning, strain hardening and time dependence 19

5 Rheological Measurement Techniques **21**
5.1 Shear rheometry 22
5.1.1 Drag Flows 22
5.1.2 Pressure Driven Flows 25
5.2 Elongational rheometry 27

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

6 Rheology and structure of functional soft matter and interfaces 30
6.1 Transient rheology 30
6.2 Superposition rheometry 31
6.3 Rheology of complex fluid-fluid interfaces 33
6.4 Optical rheometry 34

7 Further information 35

1 Introduction

Rheology is defined as the science of deformation and flow [1, 2, 3]. In principle, this definition would include everything that deals with flow, such as fluid dynamics, hydraulics, aeronautics and even solid state mechanics. However, in rheology we tend to focus on materials that have a deformation behaviour in between that of liquids and solids, i.e. with a material behavior which does not follow the classical constitutive equations of Newton (for fluids) or Hooke (for solids). Rheology can be seen as the "materials science of exceptions". The science of rheology started in the 1920s when polymers started to be produced, leading to novel polymeric substances and new colloidal fluids (e.g. paints). Classical Newtonian fluids and the Hookean elastic solids are outside the scope of rheology and material behaviour intermediate to these classical extremes will be studied in this chapter. The term *viscoelastic* is used to describe this behavior. Some fluids are however essentially inelastic, but have a viscosity which changes with the deformation rate, they are *Non-Newtonian fluids*.

Rheology is in essence an applied science, and its aim is twofold: Firstly, rheologists try to understand the interplay between structure and the flow properties [4]. This is important for the rational design and/or formulation of materials for certain applications. The measurement of rheological properties can hence be viewed as an analytical technique, where microstructural aspects are deduced from the rheological properties. Secondly, by studying the material behavior using simple deformations, fundamental relations will be derived between deformation and stress. Such equations are called "constitutive equations" or "rheological equations of state". These equations can then be used to predict the material behavior in complex deformation histories as they take place in typical process operations: e.g. extrusion, polymer film blowing, spraying, pumping etc. The measurement of rheological properties is then required to provide the adequate input, the *material functions* to be used in the constitutive models, for numerical simulations. One of the challenges we will have to address is to derive a rigorous mathematical framework which tells us exactly which material functions we need to measure.

One of the additional complications in predicting the behavior of such fluids is that the processing flows are typically complex flow fields. To decouple the complexity of the fluid mechanics from the complex material response, rheological properties are typically investigated using well defined flow fields, often called rheometric flows. The components of the stress tensor are then measured separately. Two of the most important flow fields are depicted in figure 1. The first one is a shear flow, which is the typical flow field occurring next to solid surfaces, and V is the velocity of the moving top surface, h is the height. The kinematics in a Cartesian coordinate framework are given by

$$v_x = \dot{\gamma} \cdot y \tag{1}$$

$$v_y = 0 \tag{2}$$

$$v_z = 0 \tag{3}$$

Where v_i is the velocity component in the i direction and $\dot{\gamma} = \frac{V}{h}$ is the shear rate. The shear rate has as units the inverse of time: s^{-1}. Elongational or extensional flow as in figure 1(b) occurs when there is stretching, such as at a die or nozzle exit, in film blowing, fiber spinning or flow through contractions. Controlled uniaxial elongational flows can be generated with some difficulty in the laboratory experimentally.

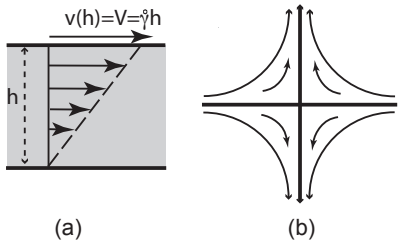


Fig. 1: *Kinematics of (a) simple shear flow and (b) an uniaxial elongational flow.*

The kinematics for uniaxial extensional flow are given by:

$$v_x = \dot{\epsilon} \cdot x \tag{4}$$

$$v_y = \frac{-\dot{\epsilon}}{2} y \tag{5}$$

$$v_z = \frac{-\dot{\epsilon}}{2} z \tag{6}$$

here $\dot{\epsilon}$ is the extensional rate. The range of deformation rates encountered during various processes spans many decades. Rheological properties are important in slow processes such as sedimentation and levelling which are dominated by shear flows ($\dot{\gamma} \sim 10^{-6}$ to 10^{-2} s^{-1}). the rheological properties of some consumer products will be tailored to impart a high viscosity at these low shear rates for a good shelf-life. Classical processes in which shear rheology further plays an important role are extrusion and injection molding in polymer processing, which occur at faster deformation rates ($\dot{\gamma} \sim 1 - 10^5$ s^{-1}), and coating flows and printing operate at very high rates ($\dot{\gamma} \sim 10^4$ to 10^6 s^{-1}) where the material response is typically non-linear. Some processes such as rubbing or chewing may not be viewed as typical processing operations, but also in those the rheological properties determine to a large extent the way these products feel. Whereas rubbing is essential shear dominated, chewing has a substantial extensional component. Fiber spinning and porous media flows also have dominant extensional flows, with the extensional rates ranging also over many orders of magnitude. However, in all aforementioned cases, the flows are still laminar. Nevertheless it has been a challenge to develop measurement techniques that can embrace this entire shear rate range (more than 8 decades)! Moreover the magnitude of the rheological properties can vary more than most other physical parameters. For example for viscosity, the SI unit is Pa s (Pascal second)[1] Working with complex fluids, viscosities can be close from to water (10^{-3} Pa /s) over polymer melts (1-10^4 Pas) to asphalt emulsions (1-10^8 Pas). Moreover, the material properties often change with rate, time and type of deformation. Measuring and describing this will be a first step to rationalising the behaviour.

[1]The cgs unit of Poise (P) is also often used. 1 poise = 0.1 Pas. Centipoises are also often encountered, as 1 cP is the viscosity of water.

2 Rheological Phenomena

2.1 Viscoelasticity

Probably the best-known example of a material with a funny behaviour is the so-called silly putty (also marketed as Thinking Putty, Bouncing Putty, Tricky Putty and Potty Putty). It was originally created by accident during research into potential rubber substitutes at GE for use by the United States military in World War II. It is a mixture of the polymer polydimethylsiloxane and borax. When this material is rolled into a small ball and dropped on a surface, it will bounce [5]. When it is left to sit on a flat surface it will flow or like any liquid will take on the shape of the container that holds it. Depending on the time scale on which the deformations take place, the material will behave like either a solid (short times) or a liquid (long times)[2]. But most of the time the material will show an intermediate behavior, displaying characteristic of a solid and a liquid at the same time. To describe this will be one of the first challenges we need to tackle. This will require a time-dependent modulus $G(t)$ or related linear viscoelastic material functions [6]. We will see that measurements in the frequency domain are particularly useful and lead to some of the most used rheological material functions, i.e; the dynamic moduli $G'(\omega)$ and $G''(\omega)$ where ω is the frequency. Experimental measurements of the linear viscoelastic properties will enable us to determine to which extent (or on which time scales) a material is fluid or solid-like, or in between and give use a dynamic fingerprint of the material in its rest state.

2.2 Non-Newtonian behaviour

Like the modulus, the viscosity is also not always a constant, it may depend on time but even more importantly on the rate of deformation. For shear flows, the shear rate dependent viscosity $\eta(\dot{\gamma})$ is the relevant non-linear material function. Many colloidal suspensions and polymer solutions show a decrease of the viscosity as the shear rate is increased, which is called *shear thinning*. Two examples are given in figure 2. An example of viscosity an shear stress data for a concentrated polymer solution, over a wide range of shear rates, is given in figure 2 (a). It is a solution of 11 wt % of a high molecular weight PIB MW_v =1.2 10^6 in pristane. This fluid is a home made version of the NIST standard reference material [7, 8]. The solution starts of at a Newtonian plateau, characterised by a zero shear viscosity, called η_0 and then shows a decrease of the viscosity over more than an order of magnitude. This is due to the stretching, orientation and disentanglement of the concentrated, high molecular weight polymer chains in solution. Shear thinning is commonly observed in many other products where shear flow is de-structuring the material.

However, flow does not always monotonically decrease the viscosity. Figure 2 (b) shows the viscosity versus shear stress for suspensions of charged polystyrene-ethylacrylate spheres with a particle radius of 125nm dispersed in water, replotted from Laun [9]. The volume fraction was varied from 0 to 0.50, temperature was kept constant at 25°C. This graph reveals the typical features that suspensions can display : At low volume fractions the viscosity of the suspensions remains essentially (pseudo-)Newtonian and the viscosity is only slightly higher than that of the dispersion medium. At higher volume fractions, the viscosity now increase dramatically,

[2]For silly putty, the relaxation time for the stress is in the range of 0.1 -1 s, the Youngs modulus (for a rapid deformation) is one to several MPa, and the viscosity for a slow deformation of order 10^5 Pas [5]

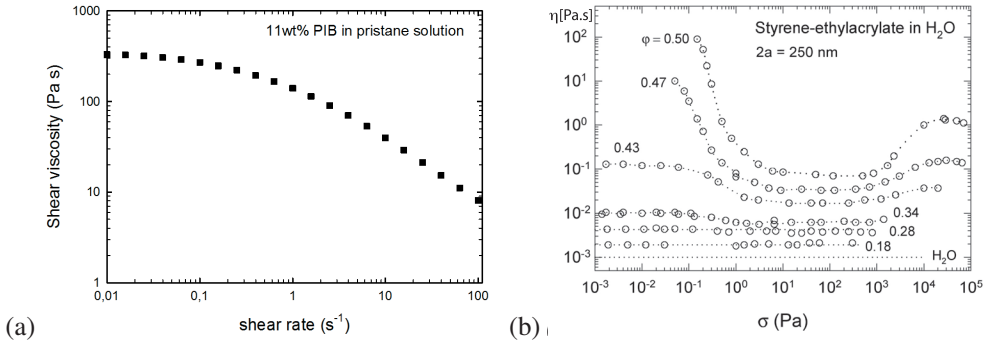


Fig. 2: *Non-Newtonian fluid behaviour (a) polymer solution containing a 11% (w/v) solution of a high molecular weight poly-(isobutene) (PIB-BASF Oppanol B200) in pristane (data set obtained by Dr. A. Vananroye, KU Leuven) (b) Viscosity shear rate curves for suspensions of styrene-acrylate latices of 125 nm radius in water, plotted for different volume fractions, from Laun [9].*

and this is accompanied by the occurrence of a complex dependence of the shear thinning and also shear-thickening is very obvious in this figure. At very high volume fractions and low shear stresses, the Newtonian plateau tends to disappear, and a yield stress starts to develop. The shear thickening becomes even more pronounced. A recent practical application of shear thickening fluids is the so-called liquid body armor, where kevlar sheets are impregnated with shear thickening fluids to create a better protection for a smaller number of Kevlar sheets [10].

2.3 Yield stress and thixotropy

Some materials apparently flow almost not up to a certain level of stress, the *yield stress* (σ_y). Plastic behaviour is common to widely different materials. However, in a well cited and provocative paper entitled "The yield stress myth", Barnes and Walters [11] however questioned the existence of the yield stress, suggesting that there is a very large Newtonian viscosity at low stresses and that the yield stress is only an apparent value, determined by the impatience of the experimentalist. Whereas it is correct to state that measuring the yield stress is difficult, the consensus is now that there are different classes of materials, such as aggregated suspensions, where a true solid-to-liquid transition occurs. Yield-stress liquids are broadly defined as materials that are solid below a critical applied stress and flow like non-Newtonian liquids at higher stresses [12, 13]. The yield stress is of particular important in a range of consumer products, as the yield stress can be used to prevent creaming or sedimentation. Concentrated dispersions often show a behaviour that can be approximated by a yield stress followed by a Newtonian behaviour. Such materials are named after E.C. Bingham who described paint in this way in 1916 and coined the word "rheology".

Yielding is often accompanied by a complex time and shear history dependence of the viscosity $\eta(\dot{\gamma}, t)$. When the time effects are reversible they are called *thixotropy* : it is defined as the continuous decrease of apparent viscosity with time under shear and the subsequent recovery of the viscosity when the flow is discontinued [14]. Thixotropy is distinct from viscoelasticity : in the latter the time scale of the material is intrinsic to the building blocks of the material, whereas

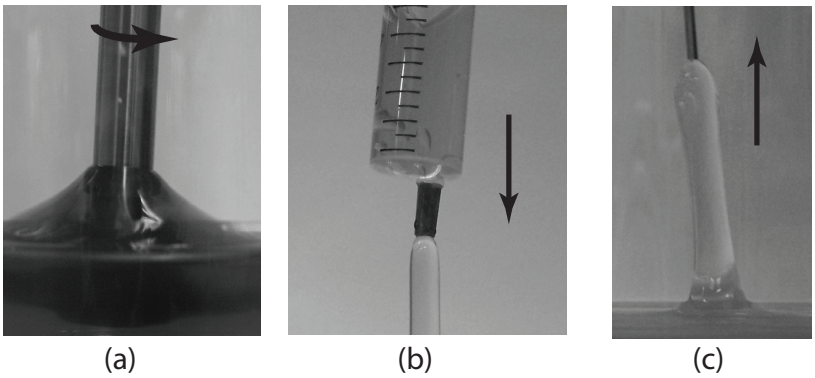


Fig. 3: *Effects of fluid elasticity and elongational viscosity for solutions of 4% polyethyleneoxide (Mw 10^6) in water: (a) Weissenberg - rod climbing effect (b) Extrudate swell (c) Ductless syphon. The arrows indicate the directions of flow.*

in thixotropy, the time scale can be set by a variable, shear history dependent microstructure, such as in an aggregating suspension.

2.4 Elastic effects

A spectacular example of elasticity in a fluid is the so-called Weissenberg or rod-climbing effect. When a rod is set to rotate in a Newtonian fluid, at high rotation rates the inertial forces push the material to the outside of the container. Exactly the opposite behaviour is observed in viscoelastic fluids, and already at much lower rotational speeds as is shown in figure 3(a). When a rod or overhead stirrer is embedded in an elastic liquid, the level of the free surface of the liquid is higher near the rod than it is at the rim of the vessel. In figure 3(a) a solution of a high molecular weight polyethyleneoxide is dissolved in water. When subject to rotation, stresses will develop along the normal axes of the flow field due to deformation and stretching of the polymer chains. In the case of a rotating rod, the stream lines are circular. Hence the normal stress in the direction of the stretching (flow) is greater in magnitude than the two mutually perpendicular components, and hence a tension in the flow direction results. These normal stresses that arise will strangle the material around the rod, pushing it up (see e.g. [2, 3]).

The phenomenon of extrudate-swell, shown in figure 3(b), is another unusual behaviour which is typical for viscoelastic fluids. For a Newtonian material the ratio of the diameter of the jet to the capillary diameter is 113% (at low Reynolds number) whereas it can be as high as a few 100% for the case of polymeric fluids and the figure reveals indeed a extradite much ticker compared to the nozzle that it exits from. Physically, the extrudate swell arises mainly because of differences in the normal stresses along and perpendicular to the stream lines in the tube. They relax as the fluid exits from the tube causing a contraction in the longitudinal direction. This is a very important phenomenon when extruding products. For example when you would make a square duct by extrusion, this swelling needs to be accounted for using a suitable die design.

In more complex flows, the non-linear material characteristics will strongly alter the flow fields themselves, due to the interplay between kinematics and dynamics. For example in contraction

flows strong vortices will appear in the entrance region of the flow, as shown very nicely in the work of Boger and coworkers [15, 16]. These vortices can even become unstable and start to swirl in the die. This is not related to turbulence, but is an instability due to the elastic nature of the fluid. It enhances the pressure drop dramatically and may lead to instabilities during processing of materials. Two often related aspects are important in entrance flow, the normal stresses and the elongational viscosity. Entrance flow problems have been a benchmark problem for numerical simulations for many decades.

2.5 Elongational viscosity

When a liquid containing large macromolecules or high aspect ratio particles is subjected to extensional deformations, the molecules and/or particles are aligned in the direction of stretching resulting in a substantial increase of the resistance to flow, and the resulting behaviour is called *strain hardening*. The elongational viscosity can be many times higher as compared to the shear viscosity, especially at high deformation rates. This may lead to phenomena as in figure 3(c), where the ductless-syphon effect is visualised. Due to the high elongational viscosity, a liquid ductless syphon is formed which allows one to draw liquid up through air, or to empty a beaker over its side. Elongational viscosity and elasticity of fluids can also be important in high Reynolds applications. For example even adding ppm levels of a high molecular weight solution can change the break-up of a jet, for example in a spray bottle or similar amounts can reduce the turbulent drag dramatically, as is sometimes used by fire fighters. A gallery of rheological phenomena is available in the beautiful book by Boger and Walters, "Rheological Phenomena in focus" [16].

3 Linear viscoelasticity

Viscous fluids and elastic solids are the two types of material behaviour classically distinguished. For elastic solids, a relation exists between the shear stress (σ) and the deformation (γ). It can be expressed as :

$$\sigma_{xy} = G \cdot \gamma \tag{7}$$

This relation is known as Hookes law. The proportionality constant G is called the rigidity or shear modulus. It has the dimensions of Pa in S.I. units. For simple liquid materials the intrinsic relation between stress and shear rate can be expressed as:

$$\sigma_{xy} = \eta \cdot \dot{\gamma} \tag{8}$$

This is known as Newtons law. The proportionality factor η is called the shear viscosity. It is expressed in Pa.s in SI units.

An important goal of rheology is to describe and characterize material behavior that is intermediate between solid and liquid behaviour. Within the general framework for describing the complex phenomena shown in the previous section, the linear regime is important because it is the building block for further work. Moreover, how complicated the non-linear regime might be, we should always find the linear viscoelastic behavior as a limiting case. Last but not least linear viscoelasticity is of practical importance: For example for linear polymer chains, measurements characterizing the linear viscoelastic properties provide insight in to how the material

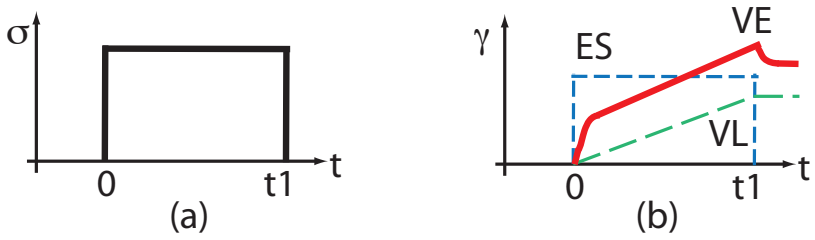


Fig. 4: *Stepwise application of a shear stress (a), response of an elastic solid (ES), a viscous liquid (VL) and a visco-elastic liquid (VE)*

will behave in processing flows, as the rate of the deformation process relative to the slowest or dominant relaxation time will determine if non-linear effects are of importance.

Viscoelastic fluids are defined as fluids which have a finite memory. The perfect solid (Hooke's law) and the perfect liquid (Newton's law) do not obey this definition. A perfect solid will remember the deformation indefinitely and return to its initial shape. A perfect liquid will stop deforming instantaneously when the stress is removed and never recover to its previous shape. All real fluids have a finite memory. For example polymeric liquids will have time scales which are on the order of 10^{-3} to a few seconds. In addition , they will often show a wide distribution of relaxation times. Alternatively one can discuss viscoelastic behaviour in terms of energy. When we apply a stress and a body is deformed, this implies that work is being performed. The perfect solid will store this energy. According to Hooke's law, the potential energy is given by :

$$E = \int \sigma_{xy} d\gamma \tag{9}$$

and this cannot be dissipated. If a perfect body would be deformed by a certain deformation γ, and subsequently it would be released, it would oscillate indefinitely (in vacuum). A perfect liquid on the other hand dissipates all the energy , to keep the fluid going I need to put energy into the fluid. Whether or not you have to take viscoelasticity in to consideration when designing a process or product will depend on the time scales of the product and the process.

3.1 Creep

Consider the following stress history (figure 4.a): at time t = 0 we apply a shear stress σ_{xy} to the material and at time t = t_1 the stress is removed. We now observe the resulting deformation of the material as a function of time. For a purely elastic solid (ES) the response is described by Hookes law. Consequently the material will deform instantaneously at t = 0, the magnitude of the deformation being σ_{xy}/G. The sample will immediately regain its original shape when the stress is released at t_1. For a viscous liquid (VL) the response is very different: as long as the stress is applied, the material will continuously flow. The shear rate is given by σ_{xy}/η. When the stress is removed at t_1 the fluid will remain stationary, neither moving forward nor backward. As the fluid does not move back to its original configuration, it could be said that it does not remember its original shape.

When the sample is viscoelastic, the initial response is as in quasi-elastic solids. Subsequently the deformation gradually increases until steady state conditions are reached. The intermediate response is referred to as delayed elasticity. When the steady state condition means that the sample continues to deform steadily, i.e. flows at a constant shear rate, it is a viscoelastic (VE) liquid. A viscoelastic solid (not shown) does not deform continuously but reaches eventually an equilibrium deformation. In order to describe creep behaviour, one has introduced the *creep compliance* $J(t)$, a time-dependent material function that is defined as:

$$J(t) = \frac{\gamma}{\sigma_{xy}} \tag{10}$$

J(t) has the dimensions of Pa^{-1}. If we restrict ourselves to the linear region, i.e. to small stresses and strains, J(t) is independent of the magnitude of the applied stress. It is then a linear viscoelastic material function. When the stress is removed suddenly at time t_1, the same regions can be identified in the viscoelastic samples as during the application of stress (figure 4). The material recoils quasi instantaneously part of the deformation. Subsequently the retarded elasticity causes the sample to recoil gradually further in time. If the sample is a viscoelastic fluid, only part of the total deformation that occurred during creep is restored. This amount is called the equilibrium elastic recoil.

The time evolution of the compliance versus time is determined by how the stress builds up in a system. The time constants which can be associated with this are the retardation times (see below). These are not always the same as the more commonly used relaxation times, but do often offer advantages in relating to the molecular or microstructure to deformation as elegantly shown by Plazek and coworkers [17, 18]. Because of the underlying linearity, closely and univocally related to each other, but the information that they emphasize is different. Creep is particularly suited for studying the long time scales, associated with the larger structural elements. Yet despite the conceptual advantages, creep measurements are relatively unpopular. In part because these are difficult to perform, in part due to ignorance of the advantages [18].

Figure 5(a) show some data obtained on the home made version of the NIST standard reference material [7, 8]. The plot in figure 5(a) shows the compliance during flow, followed by a recovery period. For low values of the stress, the compliance is independent of the applied stress level. This is the linear response regime. At larger stresses the compliance increases, and the response is non-linear.

3.2 Stress Relaxation

The visco-elastic nature of a material can also be demonstrated when a stepwise deformation is applied to the sample rather than a constant stress. The stress in a viscous liquid (VL) will drop to zero instantaneously. The elastic solid (ES) reacts to a constant strain by an instantaneous stress that remains constant in time. A viscoelastic fluid in a stress relaxation experiment reacts by gradually relaxing the stress. Hence, in this kind of experiment another time-dependent material function can be defined: the *relaxation modulus* $G(t)$, given by:

$$G(t) = \frac{\sigma(t)}{\gamma} \tag{11}$$

The stress relaxation modulus has the advantage of very naturally reflecting the viscoleastic nature of a material, and the time scale involved. The stress relaxation experiments can typically

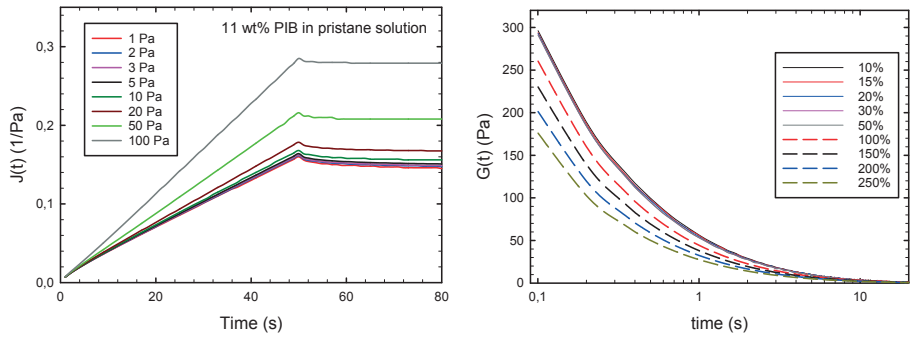


Fig. 5: *Typically observation for an entangled polymer solution (11 wt% PIB of M_{Wv} = 10^6 in pristane) (a) creep experiments at varying strains (b) stress relaxation experiments for varying step strains.*

access time scales between 0.1 (because of transducer resolutions) and a few 100 s, although faster responses ($\sim$ 20 ms) can be obtained with some effort [19]. Whereas in creep experiments, the longest retardation times are excited first, the longest relaxation times are now observable only when the stress is close to zero. The different time scales are convoluted together, and the relaxation modulus at any moment in time will be an integral over different time scales.

Figure 5(b) show some data obtained on the home made version of the NIST standard reference material but now following a step strain. For small strains, say below 100% deformation, the time dependent modulus is independent of the applied strain. At higher values of the step strain, the modulus decreases.

3.3 Small Strain Oscillatory Experiments

In these measurements, the sample is deformed sinusoidally. After a few cycles of start-up at a given frequency (ω), the stress will also oscillate with a time dependent strain amplitude γ:

$$\gamma = \gamma_0 \sin(\omega t) \tag{12}$$

where γ_0 is the strain amplitude and the rate of deformation $\dot{\gamma}$ equals:

$$\dot{\gamma} = \gamma_0 \omega \cos(\omega t) \tag{13}$$

For a viscous liquid and an elastic solid the response can be easily calculated by invoking the appropriate constitutive equations: it can be easily seen that the stress in an ES will oscillate in phase with the applied deformation, whereas for a VL the stress will be de-phased by 90° with respect to the oscillatory deformation. For a viscoelastic liquid we define a complex modulus G^* and a phase angle δ:

$$\sigma = G^* \gamma_0 \sin(\omega t + \delta) \tag{14}$$

It is convenient to analyze the stress wave into two waves, one in phase and the other one out of phase with the strain wave. The above eqn. can be written as which leads to

$$\sigma = G^* \gamma_0 \left[\sin(\omega t) \cos\delta + \cos(\omega t) \sin\delta\right] \tag{15}$$

$$\sigma = G^{'}\gamma_0 \sin(\omega t) + G^{''}\gamma_0 \cos(\omega t) \tag{16}$$

$$G^{*} = (G^{'} + i \cdot G^{''}) \tag{17}$$

The total stress can hence be viewed as the combination of a viscous and an elastic component, resulting in a complex modulus with a real elastic part, and an imaginary (viscous) part. The stress response will be dephased by an angle δ. This angle will be a measure for how elastic or how viscous my fluid is. If the fluid is viscous the phase angle will tend to 90°. When the fluid is dominantly elastic, the phase angle will tend to zero. The phase angle is hence a nice way to characterize the material on a viscoelasticity scale. The decomposit of the complex modulus leads to two moduli: G' the in-phase, elastic or storage modulus and G'' the out-of-phase, viscous or loss modulus. Hence we see the definition of viscoelasticity in terms of energetic behaviour reappearing from this experiment. The relationship between the tangens of the phase angle and the moduli is given by:

$$\tan\delta = \frac{G''}{G'} \tag{18}$$

Another way to view the experiments is in terms of a sinusoidal strain rate. This leads to the definition of a dynamic viscosity. This is also a material function and may be a more natural choice for those thinking in terms of liquids. We can define a dynamic viscosity $\eta^{'}$ for the magnitude of the viscous stress to the strain rate and an elastic part of the dynamic viscosity $\eta^{''}$.

$$\sigma = (\eta^{''} - i\eta^{'})\omega\gamma_0 \tag{19}$$

which can easily be related to the storage and loss moduli. Some examples will be given below. It is important to note that all these properties are sample size independent, they are intrinsic material properties.

When applying an oscillating stress and monitoring the corresponding strain, we obtain an oscillatory storage compliance $J'(\omega)$ and a loss compliance $J''(\omega)$. Although the complex modulus, G^* and the complex compliance J^* are readily related, this is not true for the individual components of J^* and G^* (see figure 9 below). For more information we refer to the texts by Ferry [6] or the more introductory level text by Goodwin and Hughes [24].

3.4 Linear viscoelastic modelling

Lets now try to describe the dual nature of fluids by simple mathematical models. Within this text we will limit ourselves to phenomenological models, which do not consider the molecular or structural origins of the timescales within the material at hand. What is important from this phenomological perspective, is whether or not the deformations are fast or slow, relative to the intrinsic time scales. A simple thought exercise already tells one that for systems composed of small molecules, the time scales to react to a deformation or a stress are very short. When high molecular weight compounds, supramolecular structures or colloidal microstructures are present, the relaxations will be slower and come into the time domain where most processes operate or where we as humans can observe phenomena. Molecular, microstructural or colloidal models will provide deeper insights but go beyond the scope of this introductory chapter.

We can start from simple phenomenological tools and using small shear deformations. A Hookean spring will be our model for a perfect solid. A dashpot will be our conceptual model for a perfect fluid. Maxwell (1867) first proposed a dashpot in series with an ideal spring and

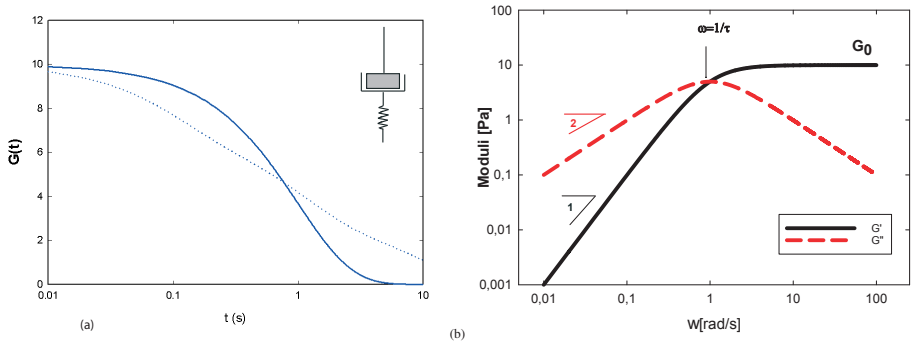


Fig. 6: *(a) Relaxation modulus for a Maxwell model (G_0 =10 Pa, τ = 1s), insert showing the spring-dasphpot (b) Storage and loss modulus for a single time Maxwell model.*

an ideal dashpot (although he did it for the viscosity of gases), as is shown schematically in the inset of figure 6(a). This is a model for a viscoelastic liquid. When we apply a deformation to this model the total deformation will be split by both elements:

$$\gamma = \gamma_1 + \gamma_2 \tag{20}$$

The stress in both elements is the same and equal to the total stress. Assuming a homogeneous deformation, the deformation rate can also be assumed to consist of a sum of the two individual motions:

$$\dot{\gamma} = \dot{\gamma}_1 + \dot{\gamma}_2 \tag{21}$$

with the stress being equal in both elements. Using Hooke's law for the ideal spring with spring constant G_0 and Newton's law for the dashpot with η_0 as the viscosity we can then rewrite the equations, leaving us with a differential equation:

$$\sigma + \left(\frac{\eta_0}{G_0}\right)\dot{\sigma} = \eta_0\dot{\gamma} \tag{22}$$

This differential equation is know as the Maxwell model. Let us first calculate the response of the Maxwell model to a step in strain. As initial conditions we impose the strain of γ_0 at times ≥ 0. We can easily calculate the stress at time zero (purely elastic response) and solving the differential equation gives use the time evolution of the stress and of the relaxation modulus which yields a decaying exponential

$$G(t) = G_0 \exp\left(\frac{-t}{\tau}\right) \text{ with } \tau = \frac{\eta_0}{G_0} \tag{23}$$

The time constant τ is the *relaxation time*. The result is plotted in figure 6 (a), and the full line on the curve gives the predictions of the above equations for a $G_0 = 10$ Pa and a relaxation time $\tau = 1s$ on a semi-logarithmic plot. In reality, a single exponential is not always observed although essentially single relaxation time fluids exist. An obvious way to generalize the Maxwell model is to allow for multiple relaxation times. This can be done by considering multiple dashpot-spring combinations in parallel. In this manner, several relaxation times are available to fit to experimental data. The dotted line is calculated using three relaxation times

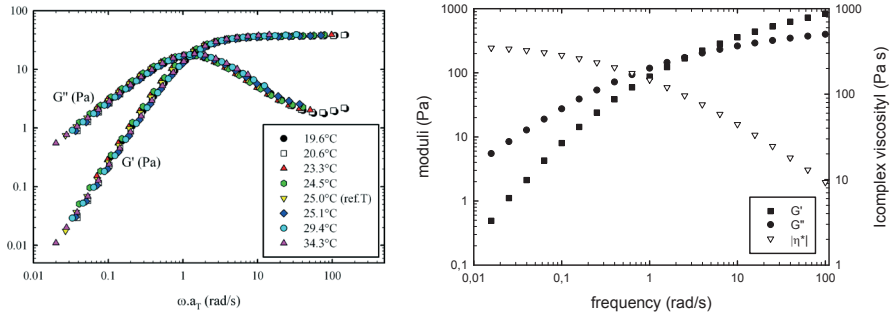


Fig. 7: *(a) Storage and loss modulus for a solution of surfactants forming wormlike micelles (100 mmol/l CpCl, 60 mmol/l NaSal, 100mmol NaCl in water) data replotted from Snijkers et al. [22] (b) Storage and loss modulus for a polymer solution containing 11% (w/v) solution of a high molecular poly-(isobutene) (PIB-BASF Oppanol B200) in pristane at 20°C.*

and shows a more gradual evolution of the relaxation modulus.

Similar to the relaxation modulus you can calculate the response of a Maxwell model to an oscillatory strain of frequency ω. It is most easily solved by using complex notation (see e.g. Ferry [6]). For a single relaxation time the result is:

$$G' = \frac{G_0 \omega^2 \tau^2}{1 + \omega^2 \tau^2} \tag{24}$$

$$G'' = \frac{G_0 \omega \tau}{1 + \omega^2 \tau^2} \tag{25}$$

The results are plotted in figure 6(b), the dotted line corresponding to the storage moduli (G') and the loss moduli (G'') being given as full lines. What you can easily see from the evolution of the moduli is that for frequencies smaller than the inverse of the relaxation time, a dominantly viscous response is observed with the loss modulus being larger than the storage modulus. This can be extended to all processes with slow motions for which indeed the dashpot or Newtonian behavior will dominate. Investigating eqn. 24-25 in the low and high frequency limits reveals some interesting and general conclusions. When deforming a viscoelastic liquid at frequencies slower than the inverse of the longest relaxation behaviour, a so-called terminal behaviour is characterised by a slope of $G^{'}$ versus ω of 2 on a log-log plot, whereas $G^{''}$ has a slope of one. At high frequencies the elastic behavior dominates and $G^{'}$ becomes independent of frequency and equal to G_0, as $G^{''}$ tends to zero. For rapidly changing stresses in general, the derivative term in the Maxwell model equation dominates and at short times this model then approaches elastic behavior. The equivalent model for a viscoelastic solid is the Kelvin-Voight model. A very good reference on linear viscoelasticity remains the monograph by Ferry [6]).

Some real materials can be characterized surprisingly well by a single relaxation time Maxwell model. For example wormlike micellar (WLM) solutions of surfactants, as can be found in many shampoos, contain flexible wormlike surfactant micelles, and the rheological properties of the solutions can be readily tuned by, e.g., concentration or temperature [20]. WLM solutions are sometimes called living systems because they combine the well-known reptation dynamics

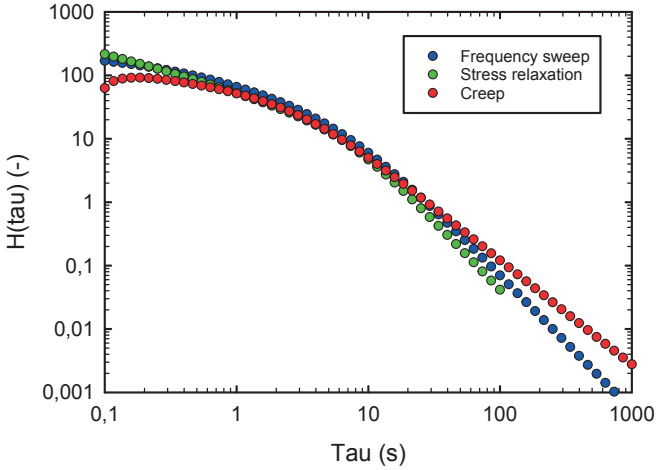


Fig. 8: *Example of relaxation spectra obtained using different linear viscoelastic measurements for a polymer solution containing 11% (w/v) of a high molecular poly-(isobutene) (PIB-BASF Oppanol B200) in pristane.*

of polymeric systems with breakup and recombination of the chains [21]. An example of such a system is shown in figure 7(a) where a solution of 100 mol/l cetylpyridiniumchloride (CpCl) in a aqueous solution containing 100 mmol NaCl and 60 mmol sodiumsalicylate (data from ref. [22]). Other types of Maxwell fluids are monodisperse emulsions and monodisperse suspensions.

Yet most complex fluids will exhibit a variety of relaxation times. For example, the polymer solution already used in figure 2 and 5, is shown in figure 7(b) [22]. The moduli display a dependence on frequency that is more smoothed out compared to the WLM. These solutions can be characterised by either a discrete set of relaxation time or by using a continuous relaxation spectrum, which for a viscoelastic fluid is typically defined as:

$$G(t) = \int_0^\infty \frac{H(\tau)}{\tau} e^{\frac{-t}{\tau}} d\tau \tag{26}$$

In this manner any arbitrary relaxation function can be described in a unique manner. This provides a very powerful tool to characterize the linear viscoelastic rheological behavior of complex fluids. The calculated relaxation spectra from the creep, SAOS and stress relaxation are given in figure 8, where you should check the relaxation time ranges which the different experiments have probed. The relaxation spectrum allows one to obtain a dynamic fingerprint reflecting the structure of complex fluids. For example for linear polymers, the relaxation spectrum can be directly linked to the molecular weight and the molecular weight distribution, although mathematically is a hard ill- posed problem, which requires adapted numerical methods. Due to the intrinsic linearity of the LV response, the different material functions are all interrelated. These relations are shown in figure 9. These relations are useful to check the internal consistency of experimental data. However, not all of these relations are numerically easy to evaluate, some are even notoriously difficult. Either open source or commercial software exists based on non-linear regularisation methods, which were introduced into rheology by Weese and

Honerkamp [23].

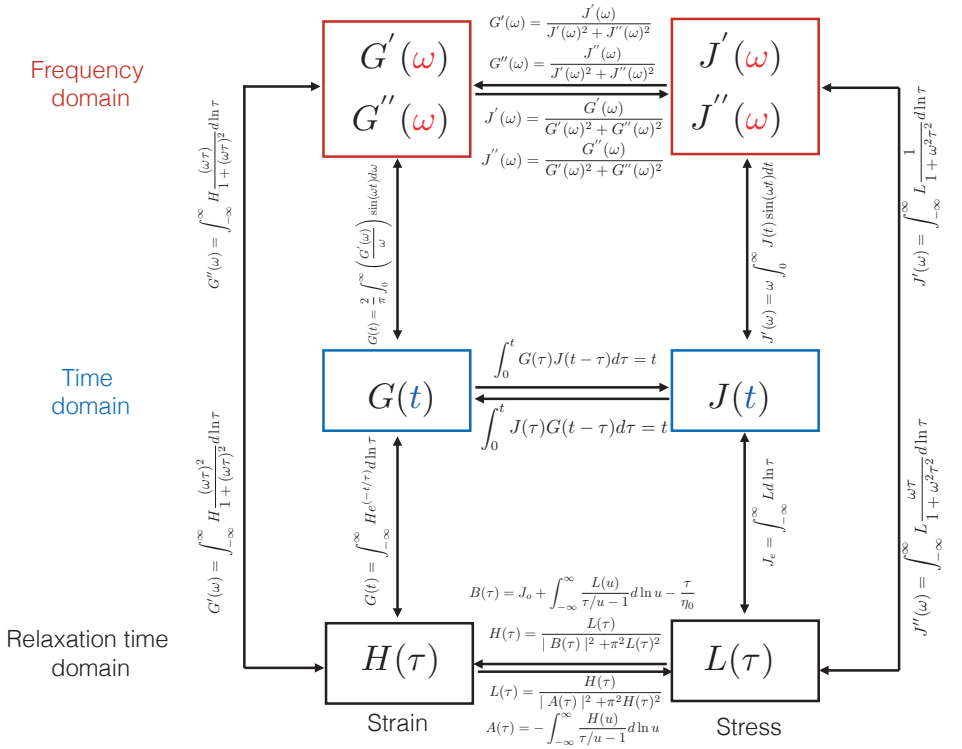


Fig. 9: *Interrelations between the viscoelastic material functions and the spectra for viscoelastic fluids.*

There are useful checks and correlations one can use. For example a useful limit to check is the relation between the zero shear viscosity and the relaxation spectrum or the relaxation modulus:

$$\eta_0 = \int_{-\infty}^{\infty} H(\tau)\tau d\ln\tau \tag{27}$$

or use

$$\eta_0 = \int_0^{\infty} G(t)dt \tag{28}$$

For example, the experimentally measured zero shear viscosity from figure 2 was 350 Pa.s. From the spectrum calculated from the relaxation modulus we obtain a value of about 250 Pa.s. The reason for this discrepancy is the finite time of the stress relaxation experiment.

Some final remarks : For many materials, the relaxation times and moduli have the same functional dependence on temperature. Relaxation times decrease strongly as temperature increases the moduli associated with relaxations are proportional to absolute temperature. This leads to the principle of time-temperature-superposition which can - for those fluids where it holds - enable one to greatly extend the measurement range, which is a principal used in obtaining the

data in figure 7. Second, the simple empirical models have been presented here for small shear deformations. To generalise the linear viscoelastic models to 3D deformations, the appropriate tensors describing relative deformations and deformations need to be used. The Maxwell model, written in appropriate tensorial form, breaks down to:

$$\boldsymbol{\sigma} + \tau \overset{\nabla}{\boldsymbol{\sigma}} = 2\eta_0 \boldsymbol{D} \tag{29}$$

where $\overset{\nabla}{\boldsymbol{\sigma}}$ is the upper convected derivative of the stress which is defined as:

$$\overset{\nabla}{\boldsymbol{\sigma}} = \frac{d\boldsymbol{\sigma}}{dt} + \boldsymbol{v} \cdot \nabla\boldsymbol{\sigma} - (\nabla\boldsymbol{v})^T \cdot \boldsymbol{\sigma} - \boldsymbol{\sigma} \cdot \nabla\boldsymbol{v} \tag{30}$$

The upper convected derivative is an extension of the material derivative, in the sense that it does not only advects with the material, but also deforms with it. The UCM model is linear in the convected reference frame only. With respect to the laboratory frame there is a non-linear response, but the material properties (relaxation time and zero shear viscosity or modulus) are constant. It is referred to as belonging to a class of quasi linear models. There is also an integral form of the LV model, called the Lodge model, see e.g. [2].

Concluding the section on linear viscoelasticity, the material functions relating to linear viscoelasticity are the relaxation modulus $G(t)$, the creep compliance $J(t)$- and the storage and loss modulus $G^{'}(\omega)$ and $G^{''}(\omega)$. For the case of linear viscoelasticity they are interrelated by the discrete relaxation times and moduli or the relaxation/retardation spectrum $H(\tau), L(\tau)$.

4 Non-Linear Viscoelasticity

The linear viscoelastic fluid presents a viscoelastic material but it will only be valid for small deformations. Two parameters are important, one being the magnitude of the deformation e.g. strain amplitude, the other being the Deborah number which is defined as the characteristic relaxation time divided by the characteristic time of the process.

4.1 Non-linear relaxation modulus

The measurements shown in figure 5(b) reveal that as the strain amplitude is increased the response becomes non-linear and the relevant material function, the relaxation modulus become dependent on strain as well as time : $G(t, \gamma)$. For most polymers or polymer solutions, the effect is simple: as the strain is increased, we find a systematic decrease of the relaxation modulus. For polymers, the intrinsic shape of the modulus/time curve remains the same, it is only shifted to smaller values on the vertical axis in a log-log plot. Hence a separate description of the time effects (the memory function) and the strain effects (a damping function) works, at least for some materials [25]. As an example we show here results from figure 5(b), by separating out the time and strain effects :

$$G(t, \gamma) = G(t)h(\gamma) \tag{31}$$

where $h(\gamma)$ is called the damping function. However, some fluids display short time strain hardening followed by strain softening starting at long relaxation times, for example associative polymers where the response is related to non-Gaussian chain stretching [26]. The relaxation modulus, nicely and physically very clearly shows how non-linear response is introduced (softening or stiffening). However, step strain experiments can be difficult, as the acceleration required needs to be fast (faster than the inverse of the relaxation times probed).

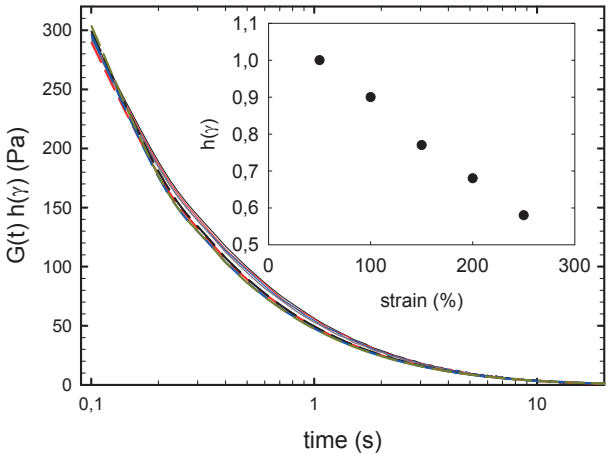


Fig. 10: *Example of rescaled modulus using a damping function for a polymer solution containing 11% (w/v) solution of a high molecular weight poly-(isobutene) (PIB-BASF Oppanol B200) in pristane, the insert shows the damming function. Legend is identical to figure 5(b)*

4.2 Large amplitude oscillatory shear flow

Also oscillatory measurements can be extended into the non-linear regime. When the strain amplitude is increased in dynamic oscillatory shear tests beyond the linear regime, interesting non-linear effects occur. This technique of large amplitude oscillatory shear flow (LAOS) has seen a revived interest in recent years, mainly because of increased computing power for signal analysis and better instrument control. As LAOS does not involve any sudden imposed jumps in speed or position, it is a relatively easy flow to generate and control. LAOS tests hence seem appealing for a broad class of complex fluids and soft matter because strain amplitude and frequency can be varied independently allowing a broad spectrum of non-linear conditions to be attained. Using techniques like fast Fourier transform, a more thorough analysis of the onset of non-linear behaviour in oscillatory measurements can be performed [27]. A detailed look at the higher harmonics offers a magnifying glass for detecting the onset of non-linear behaviour. This can be used to sensitively distinguish material properties, as e.g. polymer branching [28]. However, the deformation is complex and varies in time and the subsequent complexity of the material response makes the results more difficult to interpret in simple physical terms. Yet they have extreme sensitivity to fingerprint the non-linear response of the fluids. For a recent review on the history and topic of LAOS, we would refer the interested reader to a joint review by the leading groups in this area [29].

4.3 Normal stress differences in shear flow

More dramatic differences are observed when the strain and strain rate are increased. When a viscoelastic material is subjected to a shear flow, in addition to the shear stress (σ_{xy}), there are normal stress differences that occur. The difference between the (σ_{xx}) and the (σ_{yy}) component is called the first normal stress difference (N_1) and is the stress difference between the normal stress component on the x-plane and the normal stress component on the y-plane. The second

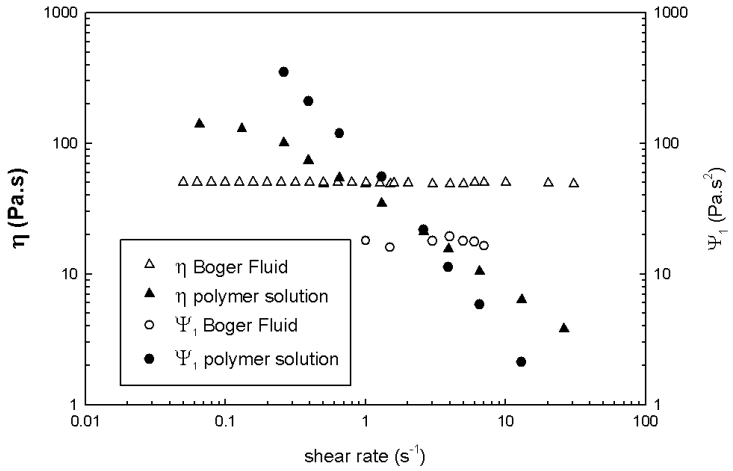


Fig. 11: *Viscosity (triangles) and first normal stress coefficients (circles) for a Boger fluid (open symbols) and an entangled polymer solution of 4% high Mw PIB in decalin (closed symbols) data from [31].*

normal stress difference is defined in a similar fashion for the normal stress components on the y-plane and z-plane. At sufficiently low shear rates, for liquids the shear stress will always become proportional to the shear rate in order to come to the limits of linear viscoelasticity. Similarly the normal stress differences (which are even functions) will become proportional to the shear rate squared. The material functions that hence can be defined are called the primary and secondary normal stress coefficients:

$$\Psi_1 = \frac{\sigma_{xx} - \sigma_{yy}}{\dot{\gamma}^2} \tag{32}$$

$$\Psi_2 = \frac{\sigma_{yy} - \sigma_{zz}}{\dot{\gamma}^2} \tag{33}$$

Figure **??** displays the viscosity and first normal stress difference for a solution of a high Mw polymer in a viscous matrix (open symbols). These fluids are quite peculiar in the sense that they have a Newtonian viscosity yet do display normal stresses differences. The shear viscosity of the system is dominated by the solvent viscosity. The elastic effects stem from the stretching of individual polymer chains leading to normal stresses quadratic in the shear rate and a constant first normal stress coefficients. These fluids have been named after David Boger, who was the first to suggest their use as model viscoelastic fluids [30]. Note that N_2 and hence Ψ_2 are much more difficult to measure, and require special equipment, as will be discussed below.

4.4 Shear thinning, strain hardening and time dependence

Most polymeric fluids will display shear thinning, rather than the behavior observed in the Boger fluids. This was already shown for the viscosity in figure 2. Yet apart from the decrease in η with shear rate there is also a decrease of the first normal stress coefficient. This is included in figure 11, where the first normal stress coefficient of a solution of a high Mw polyisobutylene

in decalin is plotted along with the viscosity. Shear thinning of the first normal stress coefficient can be observed to be even more pronounced. Shear thinning phenomena are also evident from time dependent measurements. The time dependent viscosity, which is a material function defines as:

$$\eta^+(t,\dot{\gamma}) = \frac{\sigma^+(t,\dot{\gamma})}{\dot{\gamma}} \tag{34}$$

In the non-linear region, the viscosity will show an overshoot and subsequently evolve to its steady state value. Deviations from linear viscoelasticity occur only when both the strain rate and the strain are significant. The link between linear viscoelasticity is in strict sense very small. The inverse of the longest relaxation time will indicate exactly when the strain rate will be significant. The strain limiting the LV region will do exactly the same. The relations of the LV properties to the zero shear viscosity and the low shear limit of the first normal stress coefficient are always valid, and are useful to check data consistency

$$\lim_{\dot{\gamma}\to 0} \eta(\dot{\gamma}) = \lim_{\omega\to 0} \frac{G''(\omega)}{\omega} \tag{35}$$

$$\lim_{\dot{\gamma}\to 0} \psi_1(\dot{\gamma}) = \lim_{\omega\to 0} \frac{G'(\omega)}{\omega^2} \tag{36}$$

Often, these properties exhibit similar dependencies on frequency or on the shear rate, even beyond the regime where they are expected to hold. An often used empirical correlation, known as the Cox-Merz analogy, is given by:

$$\eta(\dot{\gamma}) = \left.\frac{|G^*(\omega)|}{\omega}\right|_{\dot{\gamma}=\omega} \tag{37}$$

This analogy states that the shear viscosity is similar to the magnitude of the norm of the modulus of the complex viscosity. It is not a universally applicable used, but works relatively well for polymeric fluids, an example is given in figure12. For suspensions, deviations of this analogy are commonly observed, and can even be used to hypothesise about the nature of the flow induced structures [32].

Up to now only non-linear phenomena in shear flow have been discussed. One of the annoying things about non-linear behavior is that we cannot necessarily extrapolate the behavior in shear flow to any other type of deformation. A material function for uniaxial elongation is the time-dependent elongational viscosity

$$\eta_E^+(t,\dot{\epsilon}) = \frac{\sigma_{xx}(t) - \sigma_{yy}(t)}{\dot{\epsilon}} \tag{38}$$

for several polymer melts (branched or high Mw) the fluid will thicken, harden as the strain is increased (see below).

Remark: Non-linear viscoelastic models Non-linear constitutive modeling is a discipline by itself, which goes beyond the scope of the current text. It would require a more extensive discussion of deformation tensors and the relevant derivatives. For the current discussion it suffices to note that with a limited number of measurements of material functions, one can predict the behavior in more complex features. For example, extending the linear viscoelastic Maxwell model into the non-linear regime can be done by replacing the material derivative

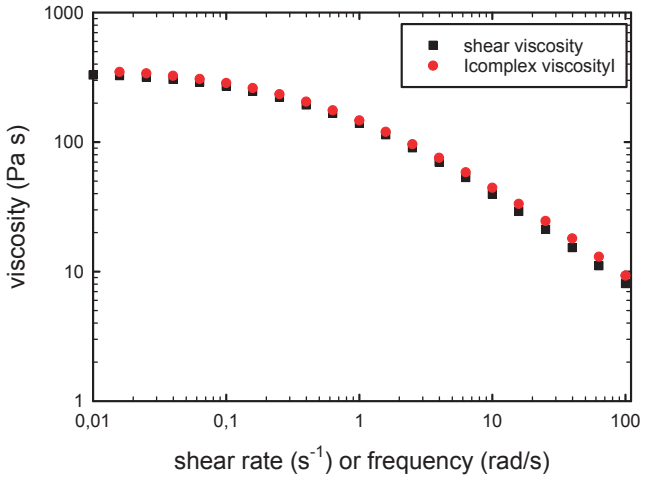


Fig. 12: *Comparison of a the non-linear steady shear viscosity and the modulus of the complex viscosity at a frequency equal to the magnitude of the shear rate for polymer solution containing 11% (w/v) solution of a high molecular poly-(isobutene) (PIB-BASF Oppanol B200) in pristane.*

in the Maxwell equation with an upper convective derivative of the stress tensor. The upper convective Maxwell model belongs to a class of quasi- linear models, i.e; they contain linear material functions but are capable of predicting non-linear response due to presence of an upper convected derivative, which is the derivative in a coordinate system embedded in the fluid. The coordinate axes are defined such that they deform with material lines. The convective term introduces non-linear terms. True non-linear models have additional terms that describe the rate at which stress builds-up that can accelerate the rate at which stress decays. To evaluate these terms we will need to have measurements of material functions characteristic of non-linear viscoelasticity. For an in depth discussion on constitutive models we refer to the monographs by Macosko [2], Morrison [3], Larson [4] for polymers, and the book by Mewis and Wagner for suspensions [32].

5 Rheological Measurement Techniques

Rheometry is for sure an underestimated discipline. Because of the complex, non-linear fluids behaviour and several intrinsic difficulties, extreme care should be taken in selecting rheological techniques and instruments. Often a single rheological measurement will be insufficient to understand the materials behaviour. The limiting relations for linear viscoelasticity should be checked at all times, as well as the independence of measurement geometry and the occurrence of time effects.

A rheometer is an instrument that measures both stress and deformation. The working equations are derived from first principles. It differs from an indexer, such as a melt flow index, because it can cary the deformation and stresses over a significant range of deformations. The working equations for rheometers and the derivations and the limits under which they are derived are

reported in detail in more elaborate texts [33, 34, 35]. A good overview of rheometrical methods and their advantages is given in the text by Macosko [2]. In the following sections we will explore the methods for shear rheology, elongational rheometry. In the next section we will end with a brief discussion on recent evolutions such as rheo-optics, interfacial rheometry and advanced rheometry.

5.1 Shear rheometry

In shear rheometry we intend to measure the small strain linear viscoelastic material function, the non-linear response observed during large strain deformations or the steady state data. The latter include the viscosity as a function of shear rate as well as the first normal stress difference. Methods for shear rheometry are distinguished by the manner in which the flow is generated: drag flows where flow is generated by moving a surface relative to one other, and pressure driven flows where pressure is applied to drive flow through a fixed geometry. The requirements set to rheometers are: flow can be analyzed and is laminar, there should be no secondary flows and boundary effects should be negligible or correctable. It is important that temperature homogeneity is ensured and that temperature control is adequate, given the strong dependency of viscoelastic functions on temperature. For large strain experiments probing the non-linear response it is best that flow is homogenous and that transducer responses are fast and that there is no inertia.

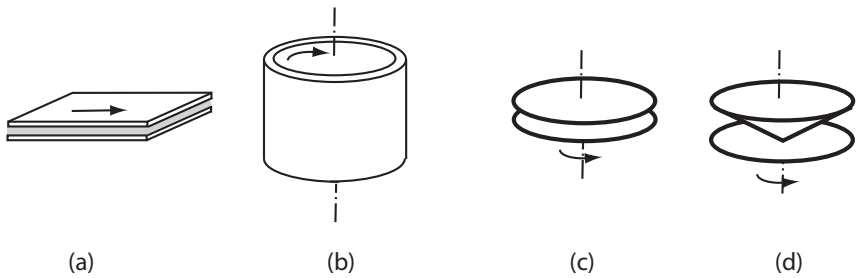


Fig. 13: *Common geometries for shear rheometry. (a) Sliding plate (b) Couette Geometry (c) plate-plate (d) Cone and plate*

5.1.1 Drag Flows

In drag flow rheometers, shear is generated between a fixed and a moving surface. The common used geometries are given in figure 13. Possibly the simplest idea is the sliding plate rheometer in Figure 13(a). The top plate is moved, a constant shear profile is imposed on the entire gap [36]. The shear rate is equal to

$$\dot{\gamma} = \frac{V_p}{h} \tag{39}$$

with V_p the velocity of the plate and h the gap width, and the dot - known as Newton's dot - represents the time derivative. Both the shear stress and normal stress difference can be directly measured. Practically this is not a very easy technique: It is difficult to keep plates parallel, especially when trying to achieve steady shear which requires large strains. The edges are open hence edge effects may play a role, their effect becoming more important as the strain increases.

Despite not being very practical, the instrument has shown to be quite useful when addressing flow in thin films and evaluating the effect of wall slip [37]. A recent evolution allows measurements in very thin, even micrometer sized gaps using the so-called flexure-based microgap rheometer [38]. The FMR has been modified to enable both measurements of normal stress differences [39] or to be combined with X-ray scattering [40], showing that this simple geometry remains appealing.

One of the most widespread rotational rheometers is the concentric cylinder geometry in figure 13(b). It goes back to the design in 1890 by Maurice Couette [41]. From the momentum balance in cylindrical coordinates, the only relevant shear stress component ($\sigma_{r\theta}$)can be found to be

$$\sigma_{r\theta} = \frac{M}{2\pi H r^2} \tag{40}$$

where M is the measured torque, H is the height of the cylinder and r is the radial coordinate. The dependence on r in eq. 41 shows that the shear stress varies with radial position. Only when the gap is small and two radii, inner (R_i) and outer (R_o), are nearly equal ($R_i/R_o > 0.99$) the shear rate can be approximated to be constant and equal to :

$$\dot{\gamma} = \frac{\Omega \bar{R}}{R_o - R_i} \tag{41}$$

with $\bar{R}$ equal to $(R_o + R_i)/2$ and Ω the angular velocity. For most (if not all) commercially available geometries this condition is not met. In this case the shear stress is not constant over the gap and a shear rate distribution will be present. The nature of the shear rate distribution will depend on the characteristics of the fluid to be measured. The amount of shear thinning will affect the velocity profile and hence the magnitude of the shear rate at the wall. this is a typical problem when dealing with non-homogeneous flows. We will discuss this in somewhat more detail below, for flow in a tube. The essential feature is that only an apparent viscosity and shear rate are being measured, and to obtain the true shear rate and viscosities, additional information is required about how the fluids properties change with shear rate. The shear rate at the inner wall can be obtained from:

$$\dot{\gamma} = \frac{2\Omega}{n\left(1 - \frac{R_i}{R_o}^{2/n}\right)} \text{ with } n = \frac{d\ln M}{d\ln \Omega} \tag{42}$$

when the change of rheological properties with shear rate is smooth. However, in the presence of a yield stress or phase transitions, even more complex shear rates distributions can be obtained, with for example only part of the field moving [42] or with so called shear-banding being present [43]. But also migration of particles in fluids can occur in these non-homogeneous flows [44], which can be amplified by the velocity profiles under the geometries, the migration leading to apparent wall slip and erroneous results.

The concentric cylinder geometry has some advantages such as a constant shear rate if $R_i/R_o > 0.99$, and due to the large measurement surface accuracy is typically good, provided end-effects are accounted for [2]. The Couette geometry is typically quite useful for settling or creaming materials. However, it is difficult to fill with high viscosity materials, and at high rates either fluid inertia or elasticity can create Taylor instabilities which entail secondary flows [45]. The concentric cylinder geometry does not allow one to obtain good normal stress measurements

(there are only some indications form rod-climbing).

Popular geometries for more viscous and elastic samples are the **parallel disks** (or plate-plate) and the **cone and plate**, shown in Figure 13(c) and (d), respectively. Their use was suggested in the 1930s by Mooney and coworkers. In the parallel disk geometry, the shear rate - given by eq. 39 - varies from being zero in the center to a maximum near the rim as the angular velocity, and hence V_p, increases with the radius. A parallel disk geometry hence entails a non-homogenous flow. The cone-and plate device overcomes this draw back, as both the velocity and the gap change with radius. Provided the cone angle is small (angle $\alpha < 0.10$ radians), the cone and plate geometry creates a homogeneous shear flow. The shear stress is constant and can be easily obtained from a momentum balance in spherical coordinates :[2, 33, 34, 35].

$$\sigma_{\theta\phi} = \frac{3M}{2\pi R^3} \tag{43}$$

where M is the torque measured and R is the radius of the cone. The shear rate is given by:

$$\dot{\gamma} = \frac{\Omega}{\alpha} \tag{44}$$

The normal stress can be obtained provided the normal force (F_z) on the plate is being measured.

$$N_1 = \frac{2F_z}{\pi R^2} \tag{45}$$

The cone and plate geometry is hence very popular, an important feature is that the flow is homogeneous. Most useful properties can be measured, both for high and low viscosity fluids. The sample volumes are small and the instruments are relatively easy to fill and clean. A cone and plate geometry is however not suited for settling materials. Care should also be taken for solvent evaporation. A stiff yet sensitive transducer is required for normal stress measurements. Most commercially available stress controlled devices have too much compliance, causing for the gap to change too much. The force rebalanced transducer technology is required, especially for transient measurements [2].

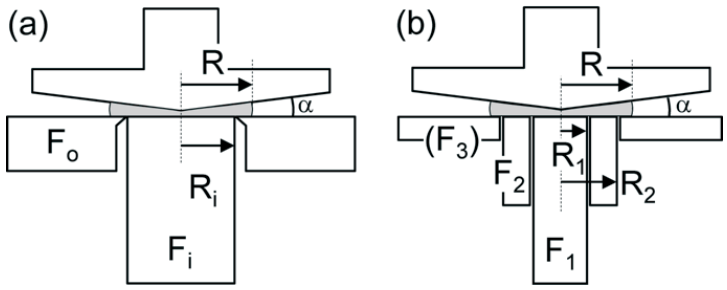


Fig. 14: *Geometries for the partitioned plate technique. (a) In this set-up, the outer partition can either be non-measuring and connected to the frame of the rheometer [47] or connected to a transducer and measure F_o [48]. (b) Partition 1 measures M and F_1, partition 2 F_2, partition 3 is a dummy to keep edge fracture away (Fig. from [47]).*

Sample loading is an important issue. The working equations for cone and plate assume spherical free surfaces. This needs to be carefully addressed and is very difficult when loading viscous

samples (such as polymers). Edge effects due to bad sample loading do contribute substantially to measurement errors. Also when centrifugal forces become important the sample will no longer be fulfilling the conditions. An instability limits the use of cone and plate geometries even more: the occurrence of shear fracture at disappointingly low shear rates. For viscoelastic samples a flow instability occurs near the free surface and the sample seems cut near the mid-plane or near the cone. This is controlled by the elasticity of the samples and the shear rate. Often shear rates of a few inverse seconds are the maximum limits.

In view of these problems with edge fracture in non-linear shear measurements and in search for a technique to measure the second normal stress difference, a variant of the cone-plate technique has been introduced 1989 by Meissner [46], the so called partitioned plate technique. In this, the plate is concentrically split in 2 or 3 partitions as shown in figure 14. The different partitions measure fractions of the total thrust, from which N_1 and N_2 can be calculated. The most appealing advantage of this technique is for viscosity measurements:

$$\eta = \frac{3M}{2\pi R_i^3 \dot{\gamma}} \tag{46}$$

The measured torque has only to be related to the innermost partition. Thus the viscosity depends on R_i, which is machined and well defined, and not on the total sample radius R, which can be off due to bad centring, edge fracture or inhomogeneous wetting of the two tool surfaces. The analytical equations for deriving N_1 and N_2 can be found in references [46, 47, 48, 49]. The main advantages of these geometries are the accuracy of the viscosity measurement, obtaining the first and second normal stress difference from a single test. The drawbacks are the increased requirements with respect to the alignment of the geometry, cleaning of the ring gap(s), which are typically 0.05-0.1 mm wide.

Parallel plate geometries are often the preferred geometry for viscous melts because of easier sample loading, the disadvantage is that the flow field is non-homogeneous. This is not a drawback for small strain functions (the response is linear), but for steady shear flows or large strains this needs to be accounted for, in manners similar to those for Couette flows. For example the shear stress, using cylindrical coordinates, is given by:

$$\sigma_{\theta z} = \frac{3M}{2\pi R^3}\left[3 + \frac{d\ln M}{d\ln \dot{\gamma}_r}\right] \tag{47}$$

Traditionally, two types of rotational instruments are distinguished, this with controlled stress and those operating under controlled strain. In the latter devices the displacement or speed (strain or strain rate) is applied to the sample and the resulting torque (stress) is measured separately by the use of a transducer. In constant stress devices a force is applied and the resulting displacement or speed is measured. Recent advances in electronics, combined with fast feedback loops, now allow for the constant stress devices to perform many of the functions of a strain controlled instrument. However, one should verify whether or not the feedback loops interfere with the non-linear response of the instruments. Reliable transient normal stress measurements can only be carried out with a controlled strain instrument with a separate transducer.

5.1.2 Pressure Driven Flows

In pressure driven flows, the wall of the flow geometry do not move but a flow is generated by applying a pressure. The first experiments of these kind were carried out, apparently independently by Hagen (in 1839) and Poisseuille (1840). The experiment is sketched in figure 15(a).

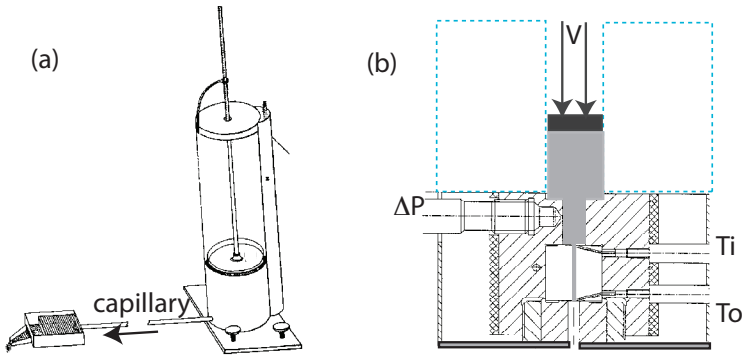


Fig. 15: *(a) Hagens experiment (adapted from Macosko [2]) (b) Detail of a current day capillary device for polymer melts, with measurement of the pressure drop and the temperatures at the in and outlet of the capillary (figure adapted from [50]*

The capillary was supplied with a test fluid from a tank which contained a float with a connecting pointer. A modern day version of the instrument is given in 15(b). Flow is driven by a plunger at constant speed (V) and the resulting pressure drop (ΔP) is measured (or vice versa). Pressure driven flows, as the ones in a capillary, entail non-homogeneous flows. In a tube, the velocity will be maximal close to the center line, where the shear rate will be small. Under the assumptions that we have steady, laminar, unidirectional, isothermal flow of an incompressible fluid, and in the absence of slip at the wall the equation of motion reduces to the following balance along the flow direction x:

$$\frac{\delta P}{\delta x} = \frac{1}{r}\frac{\delta(r\sigma_r x)}{\delta r} \tag{48}$$

From which it follows that for any fluid, the shear stress varies linearly with the radius (r) so that :

$$\sigma(r) = \sigma_w \frac{r}{R} \text{ with } \sigma_w = \frac{R}{2}\frac{\Delta P}{L} \tag{49}$$

with R the radius of the capillary, L the length and σ_w the shear stres at the wall for a pressure drop ΔP. To obtain the (non-Newtonian) viscosity, we need to relate the observable pressure drop and shear stress at the wall to the volumetric flow rate Q. The flow rate is related to the velocity v_x:

$$Q = 2\pi \int_0^R v_x r dr \tag{50}$$

Integrating this by parts and using eq. 49 to substitute dr yields:

$$\frac{Q\sigma_w^3}{\pi R^3} = -\int_0^{\sigma_w} \sigma^2 \frac{dv_x}{dr} d\sigma \tag{51}$$

Applying a differentiation with respect to σ_w using Leibnitzs rule leads to the so called Weissenberg-Rabinowitsch equation, which relates the shear rate at the wall $\dot{\gamma}_W$ tot the flow rate, taking into account the response of the fluid.

$$\dot{\gamma}_w = \left.\frac{dv_x}{dr}\right|_{\sigma_w} = \frac{4Q}{\pi R^3}\left[3 + \frac{d\ln Q}{d\ln \Delta P}\right] \tag{52}$$

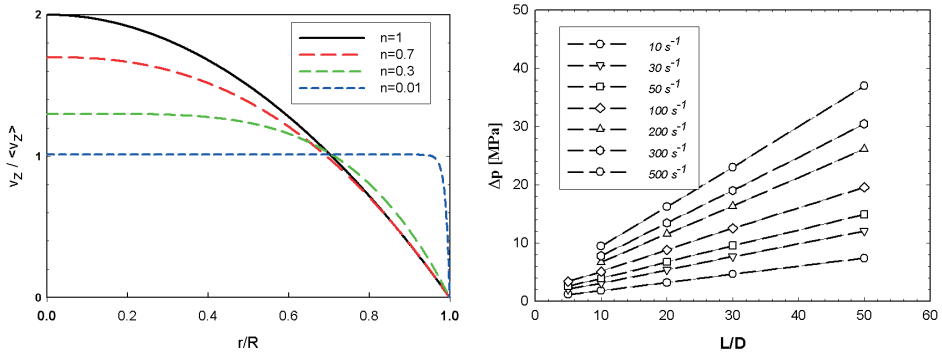


Fig. 16: *(a) Dimensionless velocity profiles as a function of the dimensionles radius for fluids with different power law indices (b) Example of a Bagley plot for a low density polyethylene at 190°C, using a capillaries of different lengths and D=1mm [50]*

The physical meaning of the Weissenberg-Rabinowitsch equation can be understood from considering the velocity profiles in a capillary. Figure 16 gives the velocity profiles for shear thinning power law fluids as a function of the power law index. For n = 1, the flow profile is parabolic,a s expected for a Newtonian material. As n decreases, the material response is more and more shear thinning. The corresponding velocity profile becomes more plug-like. The shear rate at the wall increases, or the steepness of the velocity profile changes. The shear rate is increased with respect to the Newtonian case in order to generate the same amount of shear stress at the wall.

One important remark regarding data analysis for capillary rheometry: Ideally we would like to measure $(\Delta P/L)$. However, as is shown in figure 15(b) the pressure drop is measured before the entrance of the die. The additional pressure drop occurring at the entrance of the capillary, which as we pointed out in the section 2 can be substantial due to the occurrence of large entrance vortices, needs to be accounted for. To this end, measurements with capillaries of constant diameter, but varying length can be used and plotted as in figure 16(b). Such plots are named Bagley plots [51]. The slope of the plots gives the pressure drop per unit length at a given shear rate. The entrance pressure drop can be used to estimate the elongational viscosity.

Concluding, the capillary rheometer is a simple yet accurate instrument for viscosity data. With these devices and adequate geometries, high shear rates are possible. It is a sealed system and can be pressurised (see e.g. [50]). The flow field is non-homogeneous, and the correction procedure can be laborious. Also to generate the Bagley plots, a fair amount of data is required. The upper limit of the capillary rheometry is set by the occurrence of flow instabilities either due to wall slip and the flow at the entrance of the die, which lead to shark-skin and even grossly melt fractured extrudates, respectively [52, 53].

5.2 Elongational rheometry

Most processing flows are complex superpositions of flow types and many contain substantial amounts of extensional character. The extensional viscosity is qualitatively different from

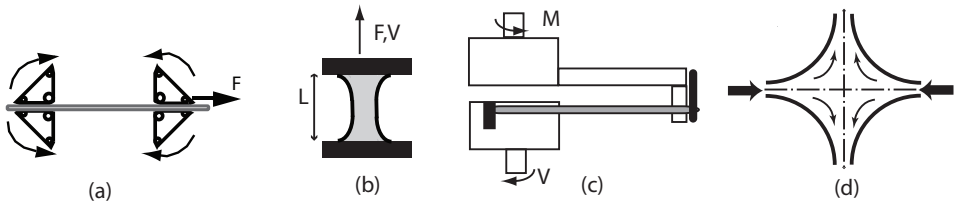


Fig. 17: *Common geometries for elongational rheometry. (a) Rotary clamp (Meissner) device (b) Filament stretching / Münstedt device / Caber setup, (c) fiber wind-up (d) opposed nozzle.*

the shear viscosity and complete characterisation should include its measurement. Extensional flows are quite difficult to realize in the laboratory.

The available techniques can be divided into methods for melts and methods for solutions. For melts devices exits which can grab onto the polymer and stretch it. One of the main techniques in polymer melt elongational rheometry has been the so-called rotary clamp technique developed by Meissner and coworkers [54] and schematically shown in figure 17(a). The sample is supported by a cushion of inert gas and, after having reached the test temperature it can be extended by clamps that make use of metal conveyor belts. The resulting tensile force is measured. For the evaluation and documentation of the test performance, a video camera records the top views of the sample that carries a marking powder to permit the evaluation of the true strain rate [54]. Nevertheless, sample preparation is not at all easy, slippage at the belts can occur and the range of deformation rates and forces is typically limited [55]. Münstedt and coworkers developed another approach whereby a polymer filament is glued to two plates, submerged in a heated oil-bath to take care of temperature homogeneity and gravity. Subsequently the holders are pulled apart either at constant deformation rate rate or stress, as is shown in figure 17(b). To achieve an elongational flow field characterized by a stretching motion in the x direction with stretching rate $\dot{\epsilon}$, the velocity fields needs to be $v_x = \dot{\epsilon}x$ and $v_r = -\dot{\epsilon}r/2$. To achieve this, the sample with initial length L_0 needs to be extended to length L with a speed the plate V_p equal to:

$$V_\mathrm{p} = \frac{dL}{dt} = \dot{\epsilon}L \text{ which entails } L = L_0 e^{\dot{\epsilon}t} \tag{53}$$

Thus the length of the sample increase exponentially. The accumulated strain $\epsilon = \dot{\epsilon}t = \ln(L/L_0)$ is called the Hencky strain. This technique has also been applied to less viscous fluids and is known there as the filament stretching technique [57]. In this technique, a small amount of sample is held between two plates, which are moved apart at an exponential rate. The liquid is held to the plates by adhesive forces. The force on one of the plates is measured as a function of time.

Recently the so called fiber wind-up technique has regained interest (figure 11(c)). This technique goes back to the 60s but was reintroduced and improved recently [58]. The device consists of two fixtures that are mounted on a rotational rheometer, the motor of which is used to stretch a filament in a controlled fashion. A filament of liquid polymer is anchored at one end in a fixture attached to the transducer, whereas the other end is wound around a second fixture that is mounted on the motor. In this manner a constant rotational speed of the motor generates a constant stretching rate in the filament. From the motor displacement the deformation of the

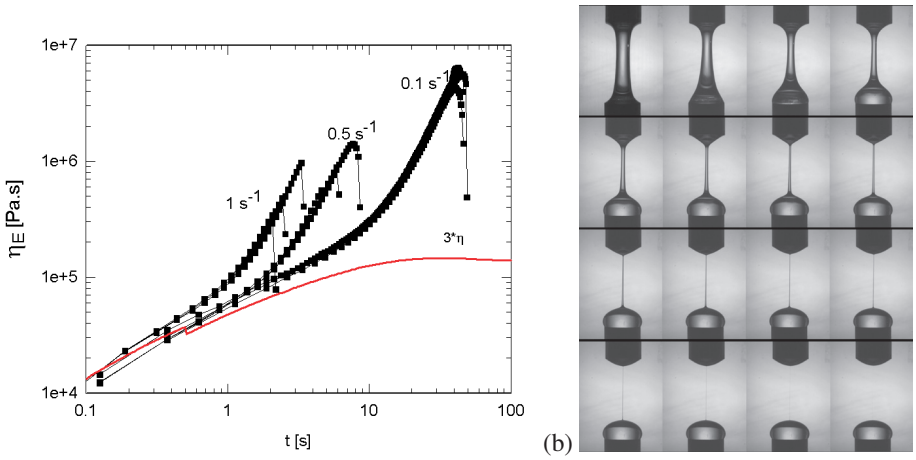


Fig. 18: *Typical data (a) polymer melts : elongational visocity as a function of time for a low density polyethylene sample at 190° in extensional flow, demonstrating strain hardening behavior. The line at $3\eta(t)$ shows the limiting behaviour in the linear limit (b) Time sequence of the capillary break-up experiment for a Boger fluid [65].*

filament can be calculated. The tension in the fiber is measured with the torque transducer. More recent designs pull symmetrically on the fibers and are commercially available based on a clever design of the fixtures [59]. Typical data taken on a commercial LDPE sample (Arkema) using the FWU technique are shown in figure 18(a). The pronounced strain hardening that can be observed makes this a suitable material for e.g. film blowing.

Solutions and other lower viscosity materials are more difficult to place in an extensional flow. The primary difficulty is that extensional flows do not have closed streamlines, which can be accomplished with simple shear flow when rotational devices are used. Since mobile fluids cannot be pulled, a number of other techniques have been developed for this purpose. For example, opposed nozzle flows create a stagnation point, where the fluid velocity is zero but the extensional gradient is finite and have been suggested as elongational rheometers [60, 61]. However, the residence time (the time that a fluid element remains in the flow field) varies from point to point. It is infinite at the stagation point and rapidly decreases away from this location, making this a difficult measurement to analyze [62]. Secondly, Contraction flows are commonly encountered in process flows and produce mixed flows that are combinations of shear and elongation. They can be analyzed to predict the extensional viscosities and rates of strains from entrance pressure drop data [63]. As a final example, a method using optical measurements of the capillary breakup of a thin fluid thread (see [65] and references therein) has recently gained interest. The rate of filament thinning and the time to breakup of a fluid element can be used, under certain conditions to measure the elongational properties of a fluid [64]. An example of a thinning of a filament is shown in figure 18(b). The mid-diameter of the filament thins under the effect of interfacial tension, and for the fluid studied (a Boger fluid: a solution of a high molecular weight polymer in a viscous solvent), it evolves in a non linear manner in time.

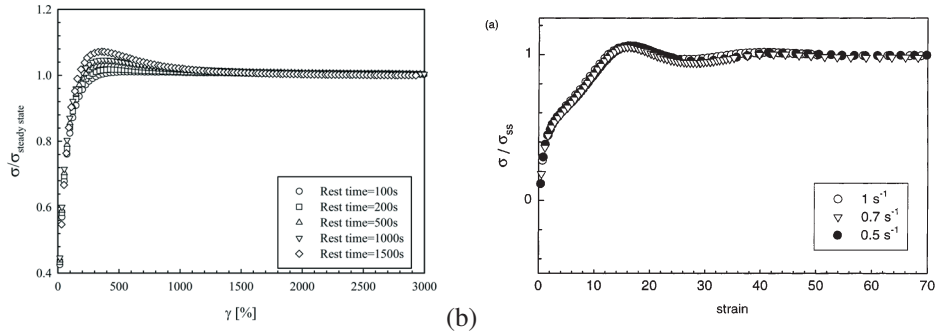


Fig. 19: *(a) Stress rescaled with respect to the stress at steady state as a function of deformation for different rest times (indicated in the legend) after 2000 s preshear at 1000 Pa in IFF experiments for a clay-polypropylene mixture at 200°C (data from [70]), (b) oscillatory stress upon flow reversal for a nematic, lyotropic liquid crystalline polymer solution (27% HPC in m-cresol) system, for different shear rates, data from [72].*

6 Rheology and structure of functional soft matter and interfaces

Whereas the classical approach of determining the above mentioned material functions in order to derive constitutive equations suffices for liquids such as homophase polymer systems, simple surfactant systems and stable colloidal dispersions, this may not be the case for more complex structured materials. For example concentrated emulsions, liquid crystalline systems, flocculated dispersions or nano-composites or other functional soft matter composites are just some of the examples that fall into a class of these structured liquids. Typical for these materials is that their microstructure spans several length scales and flow affects these possibly differently.

6.1 Transient rheology

Transient rheological measurements are very important for investigating these complex fluids. First of all, these structured materials display more pronounced transients, with the structure evolving over time scales much longer than the dominant relaxation time (τ), as obtained from linear viscoelastic measurements on the equilibrium structure. The transient rheological properties can be the essence of the rheological behaviour of the material, as in thixotropic samples as reviewed in detail by Mewis and Wagner [14, 32]. Yet the transient rheological properties often turn out to be a very sensitive probe for the changes in the structure. The methodology of using transient rheology to interrogate flow-induced changes in the microstructure has successfully been applied to several classes of microstructured functional materials.

A first example includes immiscible polymer-polymer blends, where interrupting the flow during start-up before a steady state has been reached and observing the subsequent relaxation of the shear and first normal stress difference, has proven to be a sensitive method to study the non-equilibrium structure [66, 67, 68]. Second, thixotropic suspensions or polymer-clay nanocomposites, the effects of flow on the structure are to break down the aggregate network and to possibly align the structural elements, resulting in time and shear history dependence of the

rheological properties [69, 70]. Different protocols are possible, depending on what is to be investigated. Step-stress or step-rate experiments where the sample is sheared to steady state at a given stress level and then suddenly changed, are useful to investigate the structure development during flow conditions [14]. To investigate structure development upon cessation of flow, interrupted transients can be used. For example, in interrupted forward flows (IFF), the sample is sheared until steady state is reached, upon which flow is stopped and the sample is allowed to rest. Subsequently, flow is started up again in the same direction. The increase of the stress overshoots can be used to monitor the build-up of an aggregate network. For example, figure 19 shows the evolution of the waiting time upon the stress overshoot for a clay-polypropylene nanocomposite [70]. The stress overshoot can be used to monitor the growth of a solid aggregate network upon flow cessation [69, 70, 71].

In liquid crystalline polymers, a flow reversal (FR) can be very instructive, especially in comparison with a sudden increase in shear rate as was explored in detail by Moldenaers and coworkers [73]. In this manner the effect of the orientation of the structure, existing prior to the transient, can be evaluated. For example, figure 19(b) shows the evolution of the stress in a nematic liquid crystalline polymers. The stress oscillates upon flow reversal, which moreover scales with the deformation (strain) rather than with time. This is a consequence of the fact that the response is governed by mesoscopic structure, rather than an intrinsic visoelastic time scale. For liquid crystalline polymers this corresponds to the domain structure. The length scale of the domain structure is governed by flow hence leading to a time scale set by the flow conditions, which leads to the observed strain (rather than time) scaling [4]. Measuring accurately the evolution of the normal stress differences, which typically are more sensitive to structural changes, is key to the quantitative interpretation of the observed phenomena.

6.2 Superposition rheometry

The quest for adequate measurement techniques to probe the non-linear response of complex fluids is still ongoing, as demonstrated by the recent renewed interest in LAOS [29]. It is clear that measurements in the frequency domain are appealing as the timescale of the deformation can be readily varied. However, LAOS data are only scalar in nature and the deformation history in a LAOS experiment is complex. Superposition of a small-strain oscillatory motion onto a steady or transient shear flow can provide a clearer insight into the effects of flow on the mechanisms underlying the non-linear response of rheological complex fluids [74].

In principle, dynamic moduli can also be used to probe structure during flow. It requires that a small amplitude oscillation be superimposed on the steady state flow. This can be performed with many modern stress-controlled rheometers. The oscillatory motion is then parallel to the steady state flow. In a Cartesian coordinate system, the kinematics can be written as:

$$x_1(t) = x_1(t') + [\dot{\gamma}(t-t') + \gamma_0(\sin\omega t - \sin\omega t')] \cdot x_2(t) \quad (54)$$

$$x_2(t) = x_2(t') \quad (55)$$

$$x_3(t) = x_3(t') \quad , \quad (56)$$

in which t refers to the time of observation, and t' to the history of the motion. The oscillatory motion is characterized by a peak strain γ_0 and a frequency ω. It is superposed on a steady shear flow with a shear rate $\dot{\gamma}$. in this manner the oscillatory flow is strongly coupled with the

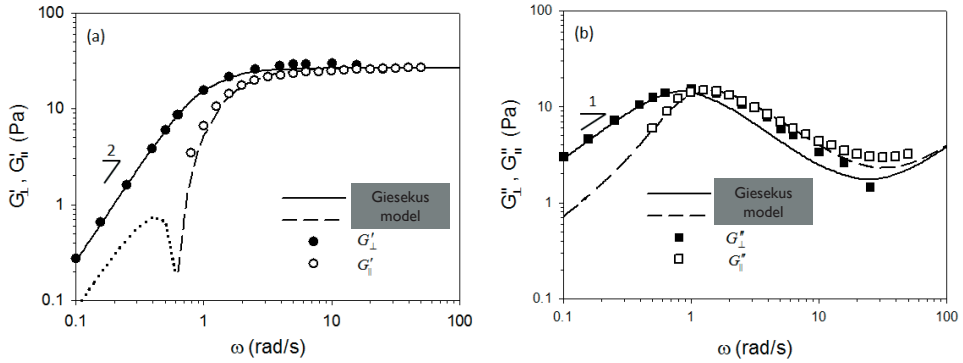


Fig. 20: *Parallel and orthogonal moduli for a wormlike micellar solution investigated containing 100 mM cetylpyridinium chloride, 60 mM sodium salicylate, and 100 mM NaCl in water. The sample is shear at $\dot{\gamma} = 0.9s^{-1}$. The Experimental values (symbols) are compared tot he results of a non-linear model Giesekus model (lines). The dotted line in a represents absolute values of negative $G''_{\parallel}$, data from Kim et al. [78].*

steady state motion and in general cannot be analysed as an ordinary oscillatory experiment without superimposed steady state flow [74, 75]. Whereas superposition moduli are in general affected by both changes in the relaxation spectra and changes in the material anisotropy, one can associate a perturbation relaxation time spectrum with each mean shear rate to illustrate what is being measured. For an integral model the prediction of the in-phase part of the moduli are given by [75]:

$$G'_{\parallel}(\omega,\dot{\gamma}) = \int_{-\infty}^{\infty}\left[H(\tau,\dot{\gamma}^2) + 2\dot{\gamma}^2\frac{\partial H(\tau,\dot{\gamma}^2)}{\partial\dot{\gamma}^2}\cdot\frac{1}{1+\omega^2\tau^2}\right]\cdot\frac{(\omega\tau)^2}{1+\omega^2\tau^2}\mathrm{d}\ln\tau \quad , \qquad (57)$$

The parallel superposition moduli not only contain the perturbation spectrum associated with each shear rate, but also the derivatives of these spectra with respect to the shear rate. The presence of this second term means that the relationships between moduli and the spectrum, expected from the theory of linear viscoelasticity no longer hold, actually even negative parallel moduli can be observed, when the second term dominates the first [75].

With a suitable designed coaxial cylinder geometry and a modern force rebalanced transducer the oscillation can also be in the transverse direction deformation and an oscillatory motion are applied at the same time but now in mutually perpendicular directions [76]. The kinematics in this case are given by:

$$x_1(t) = x_1(t') + \dot{\gamma}(t-t')\cdot x_2(t) \qquad (58)$$
$$x_2(t) = x_2(t') \qquad (59)$$
$$x_3(t) = x_3(t') + \gamma_0(\sin\omega t - \sin\omega t')\cdot x_2(t) \quad . \qquad (60)$$

Using a perturbation approach the moduli can be calculated [75]

$$G'_{\perp}(\omega,\dot{\gamma}) = \int_{-\infty}^{\infty} H(\tau,\dot{\gamma}^2)\cdot\frac{(\omega\tau)^2}{1+\omega^2\tau^2}\mathrm{d}\ln\tau \quad , \qquad (61)$$

The resulting superposition moduli are different from the parallel ones for non-linear materials, though microstructural understanding is only starting to [77, 78]. For example the non-linear properties of a wormlike micellar (WLM) solution the rate-dependent relaxation time and a rate-dependent plateau modulus can be derived. The latter provides insight into the structural anisotropy during flow at short length scales, which in this case is isotropic. Examples of superposition moduli are given in figure 20. Further analysis of the superposition moduli has been used to separate and quantify the effects of flow on the reptation and breaking of the chains [78]. The comparison of parallel and orthogonal methods can also be used, under certain conditions as a method to evaluate the flow-induced anisotropy [79], and even 2D-SAOS measurements can be performed [80]. As the methods are now commercially available, it can be anticipated that they will contribute further to the understanding of the non-linear response.

6.3 Rheology of complex fluid-fluid interfaces

Interfaces between two fluids occur just about everywhere. When the fluids are pure and simple, a single value of the surface tension suffices to characterize the interface. However, in most technological or biological applications, amphiphilic molecules such as surfactants, proteins, particles, or macromolecules will populate the interfaces. The rheological characterization of these materials at interfaces is stimulated by their use in many interface dominated materials of industries ranging from food, pharmaceutical to biomedical, as well as biological applications [81, 82]. Interfacial rheological properties play an important role in determining the stability of high interphase systems such as emulsions and foams. Interfacial rheometry is non-trivial because, in all measurement techniques the coupling of bulk and interfacial flows need to be accounted for. Additionally, unlike bulk fluids, interfaces are compressible and also dilatational deformations need to be considered.

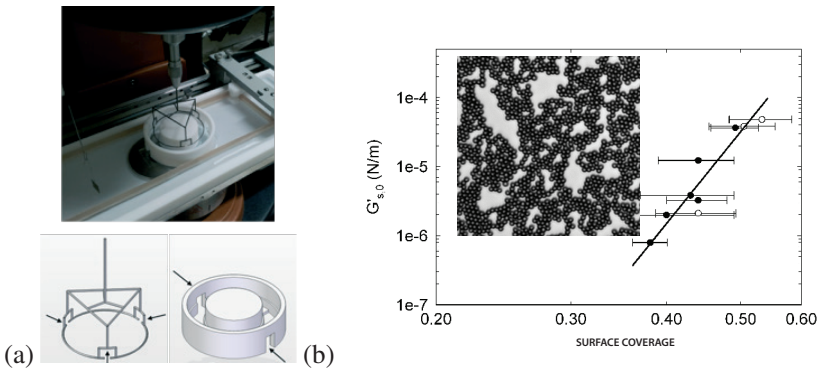


Fig. 21: *(a) Double wall ring geometry for interfacial shear rheometry (adapted from [87] (b) Plateau elastic surface modulus as a function of surface coverage for a monolayer of aggregated 3 micrometer latex particles at a water-air interface. Aggregation was induced by adding 0.4 M NaCl to the aqeous subphase), data from [90]).*

Several devices have been proposed to measure the material functions in shear, including disk or bi-cone geometries [83], and a stress rheometer where a magnetic rod is pulled at the interface [84, 85]. Figure 21(a) shows a 2D equivalent of a Double Couette geometry, using a double wall

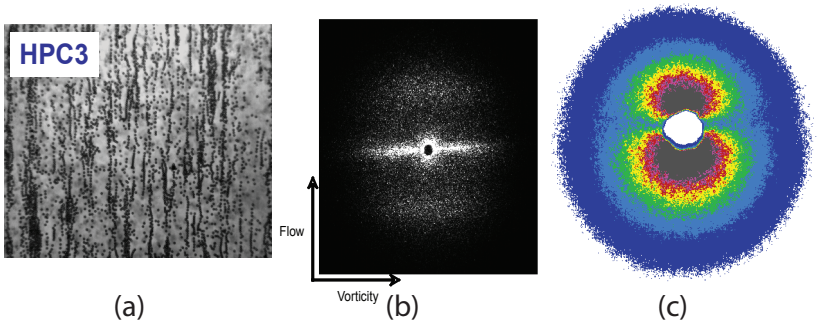


Fig. 22: *Flow induced anisotropy (a) micrograph of particle chaining of three micrometer polystyrene particles in viscoelastic fluids and (b) corresponding 2D SALS pattern (data from [31]), (c) 2D SALS pattern of a weakly aggregated suspension of nanosilica spheres (data from [95].)*

ring geometry [86]. The different linear viscoelastic material functions in shear can be readily obtained, for example figure 21(b) shows the evolution of the surface modulus as a function of surface coverage for a weakly aggregated monolayer of particles. The insert shows the structure at $\phi = 0.40$. For dilatational rheometry it is more difficult to obtain clean kinematics [82] and to separate the different material functions from the compressibility is as yet not fully accomplished [91]. Also microroheological techniques are finding their way into the field [88, 89].

The 2D systems offer an advantage that all microstructural information is also contained in one plane. This means that direct observations of the structure during rheological measurements are possible. This has been applied to colloidal crystals [92], colloidal gels [93]) and phospholipid monolayers [88], to cite a few examples.

6.4 Optical rheometry

A trend for the past 25 years has been to combine different kind of structure probing techniques with rheology. In a narrow sense, rheo-optics refers to optical methods to measure stresses. In a broader sense, rheo-optics deals with he interaction of light with flowing matter. The methods that have been applied most successfully include video-microscopy (bright field, polarized light and confocal microscopy), scattering methods (SAXS, SANS and SALS) on flowing systems, polarimetric methods providing linear birefringence and scattering dichroism. Material classes where these techniques have been extensively applied include emulsions and polymer blends, as reviewed by Tucker and Moldenaers [68] and for colloidal suspensions, as reviewed by Vermant and Solomon [94]. A key feature of the rheo-optical methods is that they are able to measure the anisotropy induced by flow. Orientation and deformation of the microstructure will cause anisotropy in the refractive index and cause it to be a tensor. Linear birefringence and scattering dichroism correspond to respectively the real and imaginary part of the refractive index tensor.

When the wavelength of the light is chosen as to avoid absorption, linear birefringence will mainly probe orientation of molecules, whereas scattering dichroism will be sensitive to larger (micrometer sized) entities such as particles or droplets. Rheo-optical methods enable very fast

and sensitive measurements, for example of particle orientation, or drop deformation. For details on techniques we refer to the monograph by Fuller [97]. When dealing with scattering or microscopy techniques, the analysis often differs from the ones discussed for equilibrium structures as the images are highly anisotropic, and often the structure is hierarchical and spans several decades in length scale (see e.g. [95, 96]). Some examples for microscopy and 2D-SALS patterns are given in figure 22, for particles in viscoelastic liquids and weakly aggregated systems. Despite the relative simplicity of these systems: the suspensions are present only in small concentrations the resulting microstructural anisotropy is significant and requires adapted analysis methods. Recent advances in microscopy have also enabled time-resolved and in-situ imaging of the structures in direct space. Fast confocal measurements and the use of super resolution microscopy provide new and exciting avenues for direct imaging of the microstructure (see the chapter by P. Lettinga in this volume and references therein). The sometimes unexpected couplings between structure and macroscopic non-linear material response, yield an exciting and very relevant area of research!

7 Further information

There are several scientific societies and journals which aim to advance the field of rheology in particular. The European Society of Rheology (ESR - http://rheology-esr.net/) is the umbrella organization for the European Societies. They organise schools and an annual conference (AERC). Membership to the ESR includes online access to the official journal of the ESR, called *Rheologica Acta* (Springer) and the journal *Applied Rheology*. The US-based Society of Rheology (SOR - http://www.rheology.org) publishes the *Journal of Rheology* and the Rheology bulletin. The SOR also organises an annual conference. Every four years, in congruence with the Olympic games, rheologists worldwide meet for the International Conference on Rheology. Other journals in the field are the *Journal of Non-Newtonian Fluid Mechanics* and the journal *Soft Matter* (RSC).

Acknowledgments

Thanks to Dr. Anja Vananroye and Naveen K. Reddy (KU Leuven) and to Dr. Frank Snijkers (CNRS, France) for the data and figures on the WLM and the PIB. Prof. Christian Clasen (KU Leuven) is acknowledged for providing the figure of capillary break up.

References

[1] H.A.Barnes, J.F.Hutton, K.Walters, *Introduction to Rheology* - Rheology Series(Elsevier Science Ltd, 1999).

[2] Ch.W. Macosko, *Rheology : Principles, Measurements, and Applications* (Wiley- VCH, 1994).

[3] F. Morrison, *Understanding Rheology* (Oxford University Press, 2001).

[4] R.G. Larson, *The Structure and Rheology of Complex Fluids* (Oxford Univ Press, 1999).

[5] R. Cross, Am. J. Phys. **80**, 870 (2012).

[6] J.D. Ferry, *Viscoelastic properties of polymers* (John Wiley and Sons, New York, 3rd edition, 1980)

[7] C. R. Schultheisz, and S. J. Leigh, NIST special publication 260-143, National Institute of Standards and Technology (2002).

[8] F. Snijkers, G.D'Avino, M. Hulsen, F. Greco, P. L. Maffettone, J. Vermant, J. Rheol., **53**, 459-480 (2009).

[9] H.M. Laun, Ang. Makromol. Chem. **123**: 335 (1984).

[10] Y.S. Lee, E.D. Wetzel, N.J. Wagner NJ, J. Mat. Sci. 38: 2825 (2003)

[11] H.A. Barnes, K. Walters, Rheol Acta **24**, 323-332 (1985)

[12] P.C.F., Moller, J. Mewis, D. Bonn, Soft Matter, **2**, 274-283 (2006).

[13] M. Denn, D. Bonn, Rheol Acta **50**:307315 (2011)

[14] J. Mewis, Wagner N.J., Adv. Coll. Int. Sci. **147-148** 214227 (2009)

[15] D.V. Boger, D.U. Hur, R.J. Binnington, J. Non-Newt. Fluid Mechanics, **20**: 31 (1986).

[16] D.V. Boger and K. Walters, *Rheological Phenomena in focus* (Elsevier Science Ltd, 1993)

[17] D.J. Plazek, J. Rheol. **36**, 16711690 (1992).

[18] D.J. Plazek, I. Echeverria, J. Rheol. **44**, 831-840 (2000).

[19] K. Dullaert, J. Mewis; J. Colloid Int. Sci. **287**: 542 (2005).

[20] H. Rehage, H. Hoffmann H Mol Phys **74**:933973 (1991).

[21] M.E. Cates, Macromolecules **20**:22892296 (1987)

[22] F. Snijkers, G. D'Avino, P.L. Maffettone, F. Greco, M.A Hulsen, J. Vermant, JNNFM **166**:363372 (2011)

[23] J Honerkamp, J Weese Rheologica Acta **32** (1), 65-73 (1993)

[24] J.W. Goodwin, R.W. Hughes, *Rheology for Chemists: An Introduction*, RSC Publishing, 2008.

[25] H.M. Laun , Rheol. Acta **17** (1): 1 (1978).

[26] L. Pellens, J. Vermant, J. Mewis, Macromolecules **38** (5): 1911 (2005).

[27] M. Wilhelm , Rheol. Acta **37**: 399 (1998).

[28] T. Neidhofer, S. Sioula, N. Hadjichristidis, M. Wilhelm, Macrom. Rapid Comm. **25** : 1921 (2004).

[29] K. Hyun, M. Wilhelm, C.O. Klein, K.S. Cho, J.G. Nam, K.H. Ahn, S.J. Lee, R.H. Ewoldt, G.H. McKinley, Rev. Poly. Sci, **36**, 1697 - 1753 (2011).

[30] D.V. Boger, J. Non-Newt. Fluid Mechanics **3** (1): 87 (1977).

[31] R. Scirocco J. Vermant, J. Mewis, J. Non-Newt. Fluid Mechanics, **117** : 183 (2004).

[32] J. Mewis, N.J. Wagner, *Colloidal Suspension Rheology*, (Cambridge University Press, 2012).

[33] K. Walters, Rheometry (Wiley & Sons, inc., New York, 1975)

[34] A.A. Collyer and D.W. Clegg (eds), Rheological Measurements, 2ed (Chapman & Hall, London, 1998).

[35] J.M. Dealy, Rheometers for molten plastics (Van Nostrand Rhein-hold, New York, 1982).

[36] A.J. Giacomin, T. Samurkas J.M. Dealy, Polym. Eng. Sci. **29**: 499 (1989).

[37] S.G. Hatzikiriakos J.M. Dealy, J. Rheol, **35**: 497 (1991).

[38] C. Clasen, B.P. Gearing, G.H. McKinley, J. Rheol. **50** (6): 883 (2006).

[39] S.J. Baik, P. Moldenaers, C. Clasen. Rev Sci Instrum. **82** (3):035121 (2011).

[40] P.Pfleiderer, S.J. Baik, Z. Zhang, G. Vleminckx, M.P. Lettinga, E. Grelet, J. Vermant, C. Clasen, Rev Sci Instrum. **85**, 065108 (2014).

[41] P. Dontula P, C.W. Macosko, L.E. Scriven, J. Rheol. **49** (4): 807 (2005) and J.M. Piau, A. Piau J. Rheol. **49** (6):1539 (2005).

[42] S.M. Fielding, Soft Matter, **3** (10): 1262 (2007).

[43] L. Becu, S. Manneville, A. Colin, Phys Rev Lett. **96**: Art. No. 138302 (2006).

[44] D. Leighton, A. Acrivos, J. Fluid. Mech. **181**: 415 (1987).

[45] R.G. Larson, S.J. Mumer, E.S.G. Shaqfeh J. Non-Newt. Fluid Mech., **51**: 195 (1994).

[46] J. Meissner, R.W. Garbella, J. Hostettler, J. Rheol. **33**(6): 843-864 (1989).

[47] F. Snijkers, D. Vlassopoulos, J. Rheol. **55**(6): 1167-1186 (2011).

[48] S. Ravindranath, S. Wang, J. Rheol. **52**(4): 957-980 (2008).

[49] T. Schweizer, W. Schmidheiny, J. Rheol. **57**(3): 841-856 (2013).

[50] A. Goubert, J. Vermant, P. Moldenaers, A Göttfert, B. Ernst, Applied Rheology, **11**: 28 (2001).

[51] E.B. Bagley, J. Appl. Phys **28**: 624 (1957).

[52] M.M. Denn, Ann. Rev. Fluid. Mech. **33**: 265 (2001).

[53] J.M. Piau, N. El Kissi, B. Tremblay J. Non-Newt. Fluid Mechanics, **34**: 145 (1990).

[54] J. Meissner, Rheol. Acta **33**: 1 (1994)

[55] T. Schweizer, Rheol. Acta **39**: 428 (2000).

[56] H. Münstedt, J. Rheol. **23**: 421 1979

[57] T. Shridar, J. Non-Newt. Fluid Mechanics, **40** : 271 (1991).

[58] M. Padmanabhan, L.J. Kasehagen, C. Macosko, J. Rheol. **40**: 473-481 (1996).

[59] M.L. Sentmanat Rheol. Acta **43**: 657 (2004).

[60] A. Keller, Prog Colloid Polym Sci **75**: 179 (1987).

[61] G.G. Fuller, C.A. Cathey, B. Hubbard, B.E. Zebrowski, J. Rheol. **31** : 235 (1987);

[62] P. Dontula, M. Pasquali, L.E. Scriven, C.W. Macosko, Rheol. Acta **36**: 429 (1997).

[63] D.M. Binding, J. Non-Newt. Fluid Mechanics, **27**: 173 (1988).

[64] J. Eggers, Rev. Modern Phys., **69**: 865 (1997).

[65] C. Clasen, J.P. Plog, W.M. Kulicke, M. Owens, C. Macosko, L.E. Scriven, M. Verani, and G.H. McKinley, J. Rheol, **50**: 849 (2006).

[66] I. Vinckier, J. Mewis, P. Moldenaers, Rheol. Acta **36**: 513 (1997).

[67] I. Vinckier, J. Mewis, P. Moldenaers., J. Rheol. **41**: 705 (1997).

[68] C.L Tucker, P. Moldenaers, Ann Rev Fluid Mech, **34**: 177 (2002).

[69] M.J. Solomon, A.S. Almusallam, K.F. Seefeldt, Macromolecules **34** (6): 1864 (2001).

[70] J. Vermant, S. Ceccia, M.K. Dolgovskij, P.L. Maffettone, Macosko C.W, J. Rheol **51** (3): 429 (2007).

[71] C. Mobuchon, P.J. Carreau, M.C. Heuzey, Rheologica Acta, **46**(8), 1045 (2007).

[72] J. Vermant, M. Mortier, P. Moldenaers , J. Mewis, Rheol Acta **38**: 537 (1999).

[73] P. Moldenaers, H.Yanase, J Mewis, ACS Symp. Series **435**: 370 (1990).

[74] R.I. Tanner, J.M. Simmons, Chem Eng Sci **22**:18031815 (1967).

[75] J. Vermant J, L.M. Walker, P. Moldenaers, J. Mewis J J Non-Newtonian Fluid Mech **79**:173189 (1998)

[76] J. Vermant, P. Moldenaers, J. Mewis, M. Ellis, R. Garritano, Rev. Sci. Instrum. **68**: 4090-4096 (1997).

[77] J.K.G. Dhont, N.J. Wagner , Phys. Rev. E. **63**: Art. No. 021406 (2001).

[78] S. Kim, J. Mewis, C. Clasen, J. Vermant, Rheol Acta **52**:727740 (2013).

[79] L.M. Walker, Vermant J, Moldenaers P, Mewis J, Rheol. Acta., **39**: 26-37 (2000).

[80] C. Mobuchon, P.J. Carreau, M.C. Heuzey, N.K. Reddy, J. J Rheol **53**:517538 (2009).

[81] R. Miller, V. B. Fainerman, J. Kragel, and G. Loglio, Curr. Opin. Colloid Interface Sci. **2**: 578 (1997).

[82] G.G. Fuller, J. Vermant, Annu. Rev. Chem. Biomol. Eng. **3**:51943 (2012).

[83] P. Erni , P. Fischer , E.J. Windhab, V. Kusnezov, H. Stettin, J. Lauger, Rev. Sci. Instrum. **74**: 4916 (2003).

[84] C.F. Brooks, G. G. Fuller, C. W. Frank, and C. R. Robertson, Langmuir **15**: 2450 (1999).

[85] S. Reynaert, C.F. Brooks, P. Moldenaers, J. Vermant, G. G. Fuller, J. Rheol. **52** : 261 (2008).

[86] S. Vandebril, A. Franck, G.G. Fuller, P. Moldenaers, J. Vermant Rheol. Acta **49**:13144 (2010)

[87] E. Hermans, J. Vermant, Soft Matter, **10**(1), 175-186 (2014)

[88] S.Q. Choi, S. Steltenkamp, J.A Zasadzinski, T.M. Squires Nat. Commun. **2**:312 (2011)

[89] J.R. Samaniuk, J. Vermant, Soft Matter. **10**(36):7023-33. (2014)

[90] S. Reynaert, P. Moldenaers, J. Vermant, PCCP, **9**: 6463 (2007).

[91] T. Verwijlen, L. Imperiali, J. Vermant; Adv. Coll. Int. Sci. **206**, 428436 (2014)

[92] E.J. Stancik, Grant T. Gavranovic, M.J. O. Widenbrant, A. T. Laschitsch, J. Vermant, G. G. Fuller; Faraday Discussions, **123**:145 (2003).

[93] K. Masschaele, J. Fransaer, J.Vermant, Soft Matter, **7** 7717-7726 (2011).

[94] J.Vermant, M.J. Solomon, J. Phys. Cond. Matt, **17**: R187 (2005).

[95] H. Hoekstra, J. Vermant, J. Mewis and T. Narayanan, Langmuir **21**:11017 (2005).

[96] J.M. Kim, A.P.R. Eberle, A.K. Gurnon, L. Porcar, Norman J. Wagner J. Rheol. (**58** 1301-1328 (2014).

[97] G.G. Fuller, *Optical Rheometry of Complex Fluids* Oxford University Press, New York , 1995.

INTERFACES



D 1 Particles at Interfaces and Membranes

T. Auth
Theoretical Soft Matter and Biophysics
Institute of Complex Systems and Institute for Advanced Simulation
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Particles at fluid interfaces **5**

3 Nanoparticles at membranes **8**

4 Biological systems **15**

5 Summary and outlook **18**

A Differential geometry and the Helfrich Hamiltonian **19**

B Triangulated membranes **24**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

A zoo of particles with different shapes and with sizes in the micrometer and the nanometer range can be fabricated using both polymeric materials and metals. Numerous applications, ranging from drug delivery [1, 2] and cancer therapy [3] to antimicrobials [4], and from viscosity modifiers for complex fluids [5, 6] to stabilizers for emulsions [7], have recently lead to an increased interest especially in nanoparticles. Particles synthesized in labs, which we mainly discuss in these lecture notes, are complemented by biological systems, such as viruses, parasites, and red blood cells that do not show less variety in size and shape than the artificial systems [8, 9, 10, 11, 12]. Suspended in bulk fluid, these small colloids with sizes of micrometers and below are strongly affected by thermal motion [13]. Structure and dynamics of the particle solution are determined by direct and depletion-mediated interactions, as well as the entropy of the particles. However, in the presence of interfaces and membranes, their interactions can be considerably altered by additional membrane-mediated and interface-mediated forces.

Non-spherical particles at interfaces can deform the interfaces in their surrounding. In order to minimize the total energy of the system and thus the total interfacial area, attractive interface-mediated interactions therefore lead to clustering of such particles. The interactions depend on particle shape and interface curvature. Identical ellipsoidal particles at a planar interface attract strongest in side-by-side orientation [14], while two unequal ellipsoidal particles can form capillary arrows [15]. Like-charged ellipsoidal particles mutually attract via capillary forces, but repell via the Coulomb interaction. Therefore, the balance of Coulomb and capillary interaction can lead to networks with many tip-to-tip contacts that are not observed for uncharged particles [16]. Cylindrical particles, however, attract in tip-to-tip orientation solely by capillary forces and form rod-like aggregates on a planar interface [17]. Particles that are studied at interfaces are often in the micrometer range and the systems can therefore be observed using light microcopy.

A single particle is bound to an interface because the decreased direct contact between the phases reduces the total energy of the system. This decrease depends both on the particle geometry and the contact angle with that the fluid interface connects to the particle. Therefore, also a single particle couples to the interface curvature: often particles are attracted to curved regions of the interface, for example around the edges of a cubic micropost [17]. Furthermore, micrometer-sized spherical particles that do not show any mutual capillary interaction on planar interfaces have been found to order on a square lattice instead of on a close-packed hexagonal lattice on a curved interface [18]. Much smaller nanoparticles at an oil-water interface can assemble to form membrane-like structures that separate two water-domains [19]. This is fundamentally different from Pickering emulsions [20], where mesoscopic particles assemble at the interface between both liquids and thereby effectively minimize the interfacial energy and stabilize the emulsion.

The interaction of particles with lipid bilayer membranes is experimentally not as well studied as the interaction of particles with interfaces. One reason might be that particle-membrane interaction is often studied for nanometer-sized particles. A direct observation of these systems using light microscopy is therefore not possible, electron-microscopy, fluorescence, and scattering techniques have to be used [21, 22, 23]. In addition, while liquid-gas or liquid-liquid interfaces are abundant, isolated homogeneous lipid-bilayer membranes that consist of identical lipids do not occur naturally. For example, biological membranes are complex membranes that consist of a variety of lipids and proteins. Furthermore, biological cells contain a cytoskeleton that can contribute actively to the interaction of particles and membranes in addition to a direct

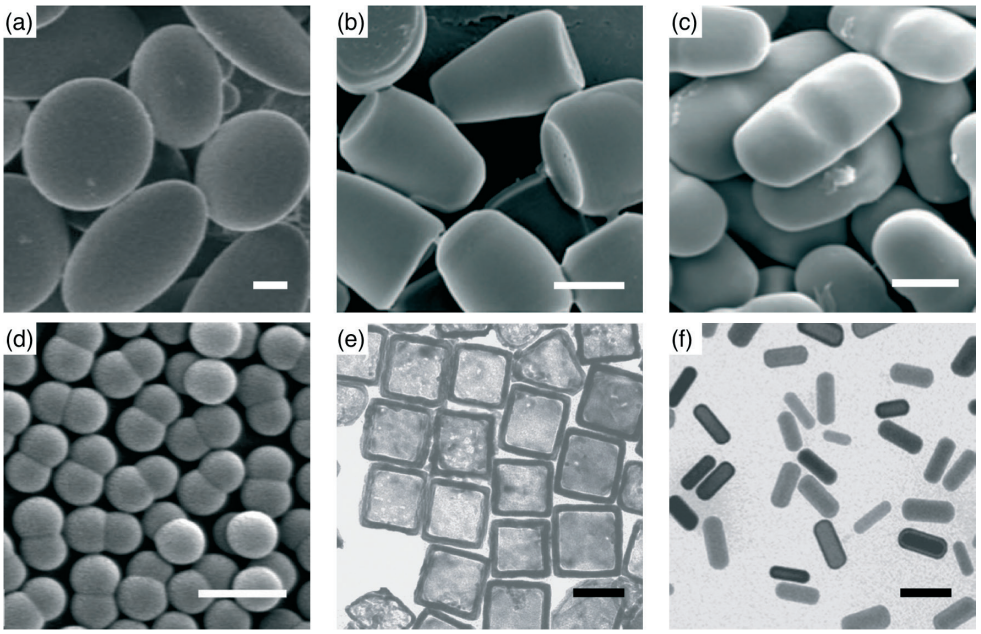


Fig. 1: *Examples for (nano-)particles: (a) oblate, disk-shaped, (b) bullet-shaped, (c) and pill-shaped polymeric particles. The length of the scale bars corresponds to* $2\,\mu\mathrm{m}$. *Taken from Ref. [30]. (d) Dumbbell-shaped polymeric particles. The length of the scale bar corresponds to* $0.5\,\mu\mathrm{m}$. *Taken from Ref. [31]. (e) Cube-like and (f) rod-like gold nanoparticles. The length of the scale bars corresponds to* $50\,\mathrm{nm}$. *Taken from Ref. [32].*

adhesive contact interaction. However, also for these more complex systems, the basic process for their interaction with particles is the particle-membrane interaction.

For particles with sizes of few nanometers, i. e. on the molecular scale, the particle-membrane interaction can be studied using molecular dynamics or Monte Carlo simulations that can have chemical specificity. Such small particles can be incorporated in the hydrophobic layer [24, 25, 26], penetrate the bilayer [27], or be wrapped by the membrane [28, 29]. However, although these simulations are coarse-grained and several atoms are treated as a single unit, the required computational effort does not allow to study nanoparticles with sizes above $20\,\mathrm{nm}$ that are usually wrapped by the membrane. In order to investigate wrapping of such larger nanoparticles and to study the cooperative interaction of many nanoparticles at a membrane, further coarse-graining of the membrane beyond the level of single lipids is required.

A successful approach to study membranes with sizes of tenths of nanometers up to several micrometers, i. e. the size of entire cells, is the Helfrich model [33]. The membrane is represented by a mathematical surface, elastic parameters characterize the mechanical properties of the bilayer. Using the Helfrich model for the membrane, a membrane and an interface can be described within the same framework. For the interface, a surface model neglects the finite interface width and the energy is characterized using an interface tension. Therefore, the total interfacial energy is proportional to the total interfacial area in the system. For a fluid membrane, the elastic constants in the surface model couple to the local membrane curvature; details of the Helfrich model are discussed in appendix A. Often the Helfrich model for a fluid

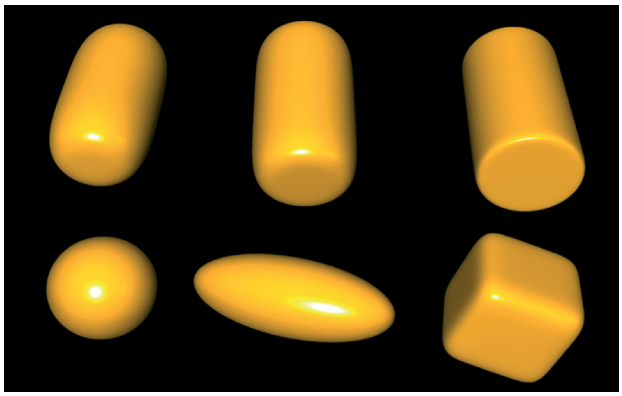


Fig. 2: *Shapes of non-spherical nanoparticles. Top row: rod-like particles of aspect ratio* 2 *with blunt tips and increasing edge curvature, defined by* $[(x^2+y^2)/a^2]^{(n/2)}+(z/b)^n=1$ *with* $n=4$*,* $n=6$*, and* $n=20$*. Bottom row: sphere, ellipsoid, Hauser's cube. We characterize the particles by their size* a*, which is the radius of the sphere, the half edge-length of the cube, and the short axis of the ellipsoidal and rod-like particles. Taken from Ref. [39].*

lipid bilayer membrane is combined with a membrane tension (that is treated analogously to the interface tension), and sometimes also with a shear modulus that characterizes the in-plane elasticity of a solid or polymerized membrane, such as a cortical cytoskeleton [34, 12, 35]. While for systems with cylindrical symmetry the energy of the surface can be calculated using shape equations, see e. g. Ref. [36], the deformation energies for more general deformations often have to be calculated using a more versatile numerical method, for example triangulated surfaces [37, 38]. Triangulated surfaces are introduced in appendix B.

Fig. 1 shows examples for particles with different shapes that can be produced in the lab. Polymeric particles of various shapes can for example be obtained by incorporating spherical particles in a thin film, liquifying the particle and deforming the film [30], dumbbell particles can be obtained using a two-step seeded emulsion polymerization [31]. A large variety of particle shapes can be described analytically using a closed expression: spherical and ellipsoidal particles, but also rod-like and cube-like particles [39], and even the egg-like shape of the malaria parasite [11]. Fig. 2 shows shapes that can be modeled as superellipsoids. Cube-like particles are described using $(x/b)^n+(y^n+z^n)/a^n=1$ with $n\geq 4$, where a and b are the sizes of the particles. For $a=b$ and $n=6$ this equation describes Hauser's cube, $x^6+y^6+z^6=a^6$. Rod-like particles are described as supereggs with $[(x^2+y^2)/a^2]^{(n/2)}+(z/b)^n=1$ and $n\geq 4$. For $n=2$, the equations for both cube-like particles and rod-like particles describe a sphere for $a=b$ and an ellipsoid for $a\neq b$. The aspect ratio of the shapes is given by b/a: the particle is prolate for $b>a$ and oblate for $a>b$. The exponent n describes the edge sharpness that increases with increasing n. For very high values of n, cube-like particles approach cuboids and rod-like particles approach cylinders.

In section 2, we discuss interface deformations around single micrometer-sized particles and the resulting interactions of particles at interfaces via capillary forces that lead to self-assembly. In section 3, wrapping of particles with fluid lipid-bilayer membranes is reviewed, in particular the interaction of single spherical, ellipsoidal, and cube-like particles with membranes. In section 3, we introduce biological systems that are related to particle-membrane systems.

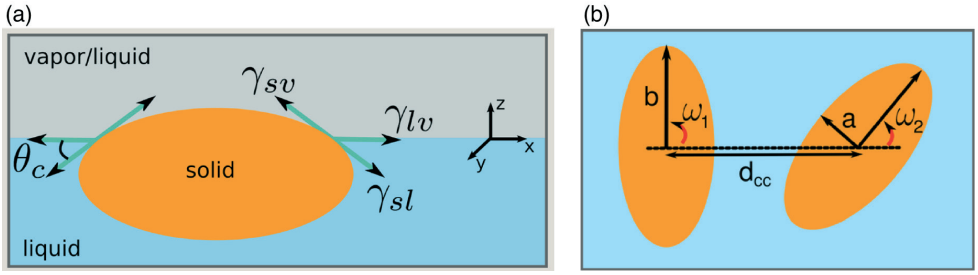


Fig. 3: *(a) Schematic representation of an ellipsoidal particle trapped at a fluid interface. The interfacial tensions between the different phases are γ_{sv}, γ_{sl}, and γ_{lv}, where s indicates the solid, l the liquid, and v the vapor (or another liquid) phase. The contact angle θ_c is determined by the Young-Dupré equation. (b) Angles ω_1 and ω_2 indicate the orientation of particle 1 and particle 2, respectively, with respect to the vector joining the centers of the two particles with aspect ratio b/a. The center-to-center distance of the particles is d_{cc}. Taken from Ref. [40].*

2 Particles at fluid interfaces

Interfaces between two liquids or between a liquid and a gas, such as an oil-water interface or an air-water interface, are characterized by their interface tensions. Typical values for interface tensions are $\gamma_{aw} = 70\,\mathrm{mN/m}$ for the air-water interface and $\gamma_{ow} = 20\,\mathrm{mN/m}$ for an ethanol/air or a methanol/air interface. Solid particles that assemble at interfaces reduce the direct contact between the phases and thereby reduce the total energy of the system. A common example for particles at interfaces are Pickering emulsions [20], mixtures of two immiscible fluids that are stabilized by particles that assemble at the interface. The maximum reduction of the interfacial energy by a single particle is given by the area that is "cut out" from the interface: e. g., for a spherical particle with radius $r = 1\,\mu\mathrm{m}$ at an oil-water interface, $\Delta E = \pi r^2 \gamma_{\mathrm{ow}} \approx 5\,10^8\,k_B T$. Particle attachment energies to an interface can therefore be of the order of $10^8 - 10^9\,k_B T$.
In general, the energy of smooth homogeneous particles at interfaces is given by

$$E = \gamma_{lv} S_{lv} + \gamma_{sv} S_{sv} + \gamma_{sl} S_{sl}\,, \tag{1}$$

where γ_{lv}, γ_{sv}, and γ_{sl} are the interface tensions. Here, s indicates the solid particle, l the liquid phase, and v the vapor phase (or another liquid phase). S_{lv} is the interfacial area between liquid and vapour, S_{sv} is the area of the particle that is exposed to the vapor, and S_{sl} the area of the particle that is in contact with the liquid. The contact angle θ_c is determined by the force balance at the contact line, see Fig. 3 (a). The particle at the interface is positioned such that it minimizes the surface energies and locally fullfills the contact angle requirement given by the Young-Dupré equation, $\gamma_{sv} = \gamma_{lv} \cos\theta_c + \gamma_{sl}$. We can therefore rewrite Eq. (1) as

$$E = \gamma_{lv}(S_{lv} - S_{sl}\cos\theta_c) + \gamma_{sv}(S_{sl} + S_{sv})\,. \tag{2}$$

The second term in Eq. (2) is independent of the wetting state of the particle, because the total surface area of the particle, $S_{sl} + S_{sv}$, is constant. For fixed contact angle, the energy of the system therefore depends only on the first term, which is proportional to the liquid-vapor interfacial tension γ_{lv} that we will refer to as γ in the following.
Micrometer-sized spherical particles at a planar interface do not deform the interface. However, to satisfy the Young-Dupré equation locally at every point on the three-phase contact line, a planar interface around an ellipsoidal particle or around a cube-like particle with rounded edges has

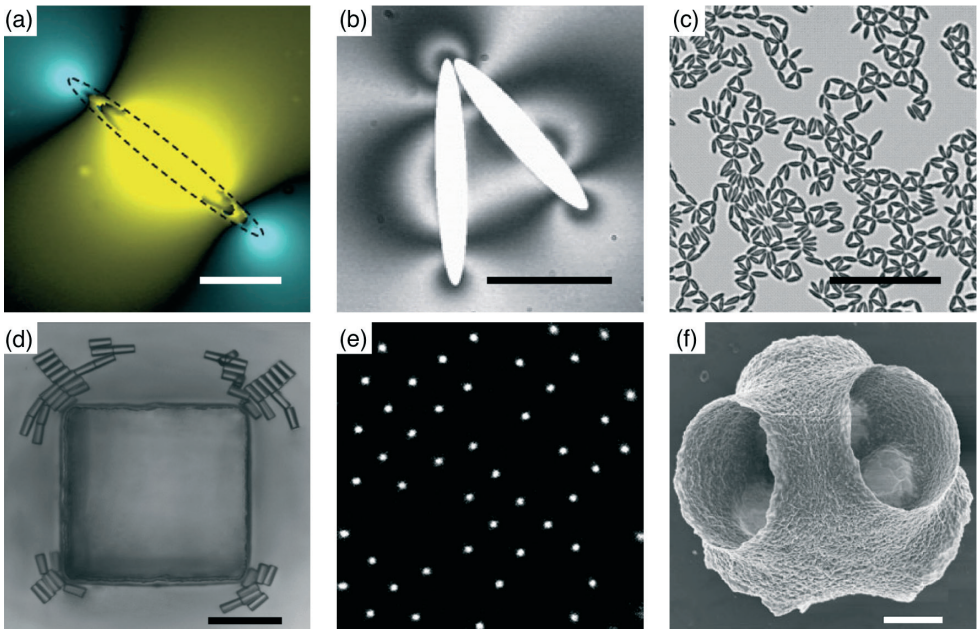


Fig. 4: *Particles at interfaces. (a) Interface deformation around a single, ellipsoidal particle. Blue color indicates that the interface is pulled down, yellow color that the interface is pulled up. The length of the scale bar corresponds to* $16.7\,\mu\mathrm{m}$. *Taken from Ref. [15]. (b) Phase-shifting-interferometry image of a capillary arrow formed by two ellipsoidal particles with different sizes. The length of the scale bar corresponds to* $12\,\mu\mathrm{m}$. *Taken from Ref. [41]. (c) Network of charged ellipsoidal particles. The length of the scale bar corresponds to* $50\,\mu\mathrm{m}$. *Taken from Ref. [16]. (d) Cylindrical particles assemble around the corners of a square micropost. The length of the scale bar corresponds to* $100\,\mu\mathrm{m}$. *Taken from Ref. [17]. (e) Spherical particles with fluorescent core on a deformed interface. The particle density is* $\approx 0.2\mu\mathrm{m}^{-2}$. *Taken from Ref. [18]. (f) Colloidosome, stabilized by 15-nm hydrophobic* SiO_2 *nanoparticles dispersed in an oil phase and adsorbed to the two water/oil interfaces. The length of the scale bar corresponds to* $10\,\mu\mathrm{m}$. *Taken from Ref. [19].*

to deform for $\theta_c \neq 90°$ [40]. Fig. 4 (a) shows the interface deformation around a single, prolate ellipsoidal particle for a contact angle $\theta_c < 90\,°$. The interface is pulled down at the tips of the ellipsoid and pulled up at the sides of the ellipsoid. Because the energy of the system increases with the interfacial area, the overall tendency of a system containing many particles is to reduce its total interfacial area. The overlap of the interface deformations therefore leads to mutual long-range interface-mediated interactions between the particles, also called (lateral) capillary forces, see e. g. Refs. [14, 15, 17]. Fig. 4 (b) shows self-assembly of two different ellipsoids as a capillary arrow. In general, capillary forces determine the interaction of particles at interfaces together with direct interactions, such as electrostatic, magnetic, and elastic interactions [42, 43, 44]. Fig. 4 (c) shows a network formed by charged ellipsoidal particles; the side-by-side contacts that are favourable for networks of ellipsoidal particles formed by capillary interaction only are partially replaced by tip-to-tip contacts due to the balance of electrostatic repulsion and capillary attraction.

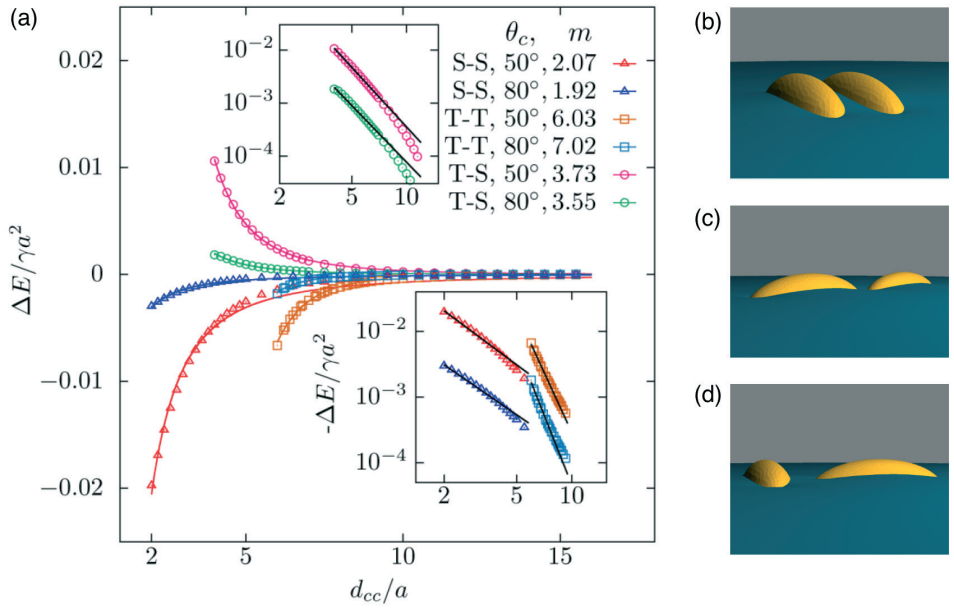


Fig. 5: *(a) Interaction energies $\Delta E/\gamma a^2$ for two identical ellipsoidal particles with aspect ratio $b/a = 3$ and contact angles θ_c = 50° and $\theta_c = 80°$. The energies are plotted as function of d_{cc}/a for side-by-side (S-S), tip-to-tip (T-T), and tip-to-side (T-S) orientation. Inset: fit of the numerical data using $\Delta E/\gamma a^2 = k(d_{cc}/a)^{-m}$; the fit parameters are given in the figure and tabulated in the supporting information. (b)-(d) Two ellipsoidal particles with aspect ratios $b/a = 3$ for contact angle $\theta_c = 50°$ in side-by-side, tip-to-tip and tip-to-side orientation. Taken from Ref. [40].*

The importance of gravitational forces can be characterized by the Bond number, also known as Eötvös number Eo, which is the ratio of gravitational and interfacial forces, $\mathrm{Bo} = \Delta\rho g a^2/\gamma$. $\Delta\rho$ is the density difference between particle and fluid, g is the gravitational constant, and a characterizes the size of the particle. Gravitational forces can be neglected if the Bond number is much smaller than 1. For a typical density difference between silica and water and for particle sizes in the micrometer range [18], $\mathrm{Bo} \approx 10^{-8}$. Hydrodynamic forces can be neglected if the capillary number, $\mathrm{Ca} = \eta v/\gamma$, is much smaller than 1. The capillary number is the ratio of viscous forces and capillary forces, where η is the viscosity of the fluid and v is the velocity of the particle. For typical velocities of $100\,\mu\mathrm{m/s}$ for particles with sizes of few micrometers that attract via capillary interactions in water [45], $\mathrm{Ca} < 10^{-5}$. Therefore, both direct hydrodynamic interactions and a distortion of the interface due to particle motion do not have to be taken into account for micrometer and submicrometer-sized particles.

The interaction energies between two particles are expressed with respect to the undisturbed flat interface in absence of the particles, such that the energy of two non-interacting particles (at large distances) vanishes. Lengths are usually given in units of the particle size a, energies as $\Delta E/\gamma a^2$. For a typical particle size $a \approx 1\,\mu m$ (Fig. 2) and a typical interfacial tension $\gamma \approx 20\,\mathrm{k_B T/nm^2}$, $\gamma a^2 \approx 2 \times 10^7\, k_B T$. In Fig. 5, the capillary interaction energy of two ellipsoidal particles is plotted for side-by-side, tip-to-tip, and tip-to-side orientation for two

different contact angles as function of the center-to-center distance d_{cc}. The particles attract each other in side-by-side and tip-to-tip orientation and repel each other in tip-to-side orientation. The side-by-side orientation is energetically most stable. The interaction potential for two non-spherical particles in the near field can be fit by effective power laws. We fit the power law $\Delta E/\gamma a^2 = k(d_{cc}/a)^{-m}$ to the data with exponents $m \approx 2$ for side-by-side orientation, $m \approx 6-7$ for tip-to-tip orientation, and $m \approx 3.5-4$ for tip-to-side orientation. For $\gamma \approx 20\,\mathrm{k_B T/nm^2}$, typical bond energies for two particles in side-by-side orientation are $6\,10^5\,k_BT$ and $4\,10^6\,k_BT$ for $\theta_c = 80°$ and $\theta_c = 50°$, respectively. These bond energies are therefore much larger than thermal energies, such that network structures that are experimentally observed for particles at interfaces can be metastable.

While interaction energies for two ellipsoidal particles in the near field have to be calculated numerically, e. g., using triangulated surfaces, interaction energies in the far field can be calculated analytically using a quadrupolar approximation. The interaction potential for two particles at large distances is [14]

$$\frac{\Delta E_{cap}^{quad}}{\gamma a^2} = -3\pi \cos(2\omega_1 + 2\omega_2) \left(\frac{\Delta u}{a}\right)^2 \left(\frac{d_{cc}}{a}\right)^{-4}, \tag{3}$$

where d_{cc} is the distance between the centers of mass of the particles. The angles ω_1 and ω_2 describe the particle orientation with respect to the line that joins the centers of the particles for particle 1 and particle 2, respectively, see Fig. 3 (b). For $\omega_1 = \omega_2 = 90°$, the particles are oriented side-by-side (S-S) and for $\omega_1 = \omega_2 = 0°$, the particles are oriented tip-to-tip (T-T), see Figs. 5 (b) and (c). Eq. (3) predicts attraction for S-S and T-T, but repulsion for tip-to-side (T-S) orientation with $\omega_1 = 0°$ and $\omega_2 = 90°$, see Fig. 5 (d). The interaction potential decays as $(d_{cc}/a)^{-4}$, which differs considerably for the power-law exponents that have been extracted from the fits in the near field. Interestingly, the magnitude of the interaction energies predicted by Eq. (3) are equal in S-S, T-T, and T-S orientation for equal center-of-mass distances.

Fig. 4 shows further examples for particles at interfaces that are mentioned for completeness, but will not be discussed in these lecture notes. Curved interfaces provide a potential landscape for particles. For example, cylindrical particles that mutually attract in tip-to-tip orientation form rod-like aggregates that are attracted to the deformed interface around the edges of a square micropost, see Fig. 4 (d). Spherical particles that do not interact on a planar interface due to the lack of interface deformation in their surrounding have been observed to order on a square lattice instead of the close-packed hexagonal lattice on a curved interface, see Fig. 4 (e). With a double emulsion technique and 15-nm hydrophobic nanoparticles dispersed in a thin film of the oil phase, "colloidal membranes" can be formed, where the particles assemble at and stabilize the oil-water interfaces. These colloidosomes can have various shapes, e. g. the non-spherical shape shown in Fig. 4 (f).

3 Nanoparticles at membranes

Nanoparticles are particles with an extension smaller than $100\,\mathrm{nm}$ in at least one dimension. Examples for incorporation of nanoparticles into a lipid bilayer, penetration of a lipid bilayer, and wrapping by a lipid bilayer are shown in Fig. 6. In this section, we only consider wrapping of nanoparticles and we use the Helfrich model for the lipid bilayer membrane. The mechanical properties of the membrane are characterised by the bending rigidity κ and the Gaussian saddle-splay modulus $\bar{\kappa}$ that couple to the eigenvalues of the curvature tensor, mean curvature and

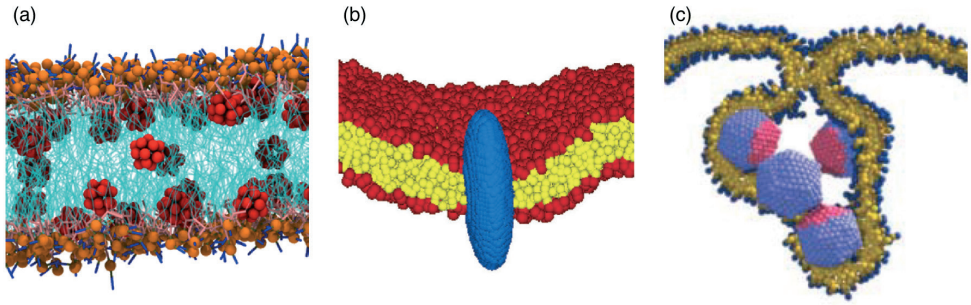


Fig. 6: *Interaction of nanoparticles with membranes. (a) Incorporation. Fullerenes in a POPC membrane. Taken from Ref. [25]. (b) Penetration. An ellipsoidal nanoparticle penetrates a lipid bilayer. Taken from Ref. [27]. (c) Wrapping. Larger nanoparticles get wrapped by lipid bilayer membranes. Taken from Ref. [29].*

Gaussian curvature, respectively. The mean curvature is the arithmetic mean of the minimal and the maximal curvature at each point of the membrane, $H = (c_1 + c_2)/2$, while the Gaussian curvature is the product of these two principal curvatures, $K = c_1 c_2$. In addition, a finite spontaneous curvature c_0 can model a tendency of the membrane to be curved, for example due to an asymmetry of the lipid monolayers in biological membranes. The energy of the membrane is given by the Helfrich Hamiltonian [33],

$$\mathcal{H} = \int_S dS \, \left[2\kappa(H - c_0)^2 + \bar{\kappa} K\right] , \tag{4}$$

where the surface integral is calculated over the entire membrane area, see also appendix A. For a homogeneous closed membrane and if the topology does not change, $\int dS\, K = \text{const.}$, therefore the Gaussian curvature does not affect the energy for shape changes; the corresponding energetic contribution is thus often neglected.
Eq. (4) describes the deformation-energy cost for a tensionless lipid bilayer membrane. In addition, we introduce a membrane tension that might originate from a cytoskeleton. Furthermore, in the following we consider a symmetric lipid bilayer in a symmetric environment, thus $c_0 = 0$. In order to wrap a nanoparticle, an energy gain that favours wrapping is required. We therefore introduce a contact energy for membrane and particle that models any homogeneous adhesive interaction. The total energy for the particle-membrane system is

$$\mathcal{E}_{\text{tot}} = \int_S dS \left[2\kappa H^2 + \sigma\right] - w \int_{S_{\text{ad}}} dS \, , \tag{5}$$

where S is the entire membrane area, S_{ad} the adhered membrane area, σ the membrane tension, and w the adhesion strength for the interaction between membrane and nanoparticle. This continuum model is applicable for particle sizes that are larger than a few times the thickness of a lipid bilayer, which is about $5\,\text{nm}$. For particle sizes smaller than $\sqrt{\kappa/\sigma}$ both bending energy and surface tension contribute, for larger particle sizes surface tension is dominant [46].
For convenience, we describe our system using the particle size a (Fig. 2) as characteristic length scale, typically $20 - 100\,\text{nm}$, and the bilayer bending rigidity κ as characteristic energy scale, typically $10 - 100\,\text{k}_\text{B}\text{T}$. We then define dimensionless parameters, such as the reduced

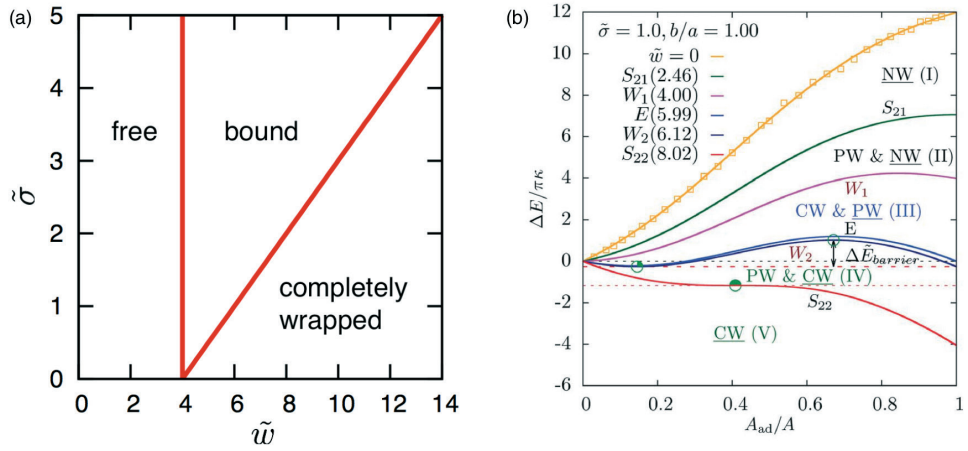


Fig. 7: *(a) Wrapping phase diagram for a spherical particle at a tensionless membrane. (b) Energies for wrapping a spherical particle as function of the wrapping fraction $A_{\rm ad}/A$ for reduced membrane tension $\tilde{\sigma} = 1$. The figure shows the wrapping energy profiles for six adhesion strengths: the numerically calculated data for zero adhesion strength and the corresponding fit function, E with equal energy for the non-wrapped and the completely-wrapped states ($\tilde{w} = 5.99$), the binding transition W_1 between the unwrapped and the partially-wrapped ($\tilde{w} = 4.00$), the binodal W_2 between the partially-wrapped and the completely-wrapped state ($\tilde{w} = 6.12$), and the spinodals S_{21} and S_{22} that are associated with W_2 (for $\tilde{w} = 2.46$ and $\tilde{w} = 8.02$ respectively). Phase boundaries separate 5 regimes in the phase diagram with stable and metastable completely wrapped (CW), partially-wrapped (PW), and and non-wrapped states (NW); the stable state is underlined. The energy minimum for the partially-wrapped state, the energy maximum for the barrier, and the inflection point for the spinodal are marked by circles. Taken from Ref. [47].*

deformation energy $\tilde{E} = \mathcal{E}_{\rm tot}/(\pi\kappa)$ and the reduced membrane tension $\tilde{\sigma} = \sigma a^2/\kappa$. The reduced adhesion strength is defined as $\bar{w} = wA/2\pi\kappa$, where A is the particle surface area, and as $\tilde{w} = 2wa^2/\kappa$. Using these dimensionless parameters instead of absolute values, our predictions apply for arbitrary particle sizes; the definitions of these parameters are consistent with those in Refs. [46, 48]. The Hamiltonian in Eq. (5) reads

$$\tilde{E} = \frac{1}{2\pi a^2}\left(\int_S dS\,\left[4(aH)^2 + 2\tilde{\sigma}\right] - \tilde{w}\int_{S_{\rm ad}} dS\right). \tag{6}$$

If the energy of the free membrane around the particle is neglected, the deformation energy for wrapping a spherical particle of radius a can be calculated analytically. Because of the homogeneous mean curvature of the sphere $H_{\rm sph} = a^{-1}$, the deformation energy density is $2\kappa/a^2$. For the sphere to be wrapped, the adhesion energy gain needs to exceed the deformation energy costs, therefore $w > 2\kappa/a^2$. If there is a membrane tension, also the tension energy has to be overcome and for complete wrapping, $w > \sigma + 2\kappa/a^2$. However, it can be shown that binding of the particle is independent of the membrane tension [46]. In reduced units, the binding transition is therefore $\tilde{w} = 4$, while the transition to complete wrapping is given by $\tilde{w} = 4 + 2\tilde{\sigma}$. Based on the discussion above, we can construct a wrapping phase diagram for a spherical particle, see Fig. 7 (a). An infinite, tensionless membrane around a partially-wrapped

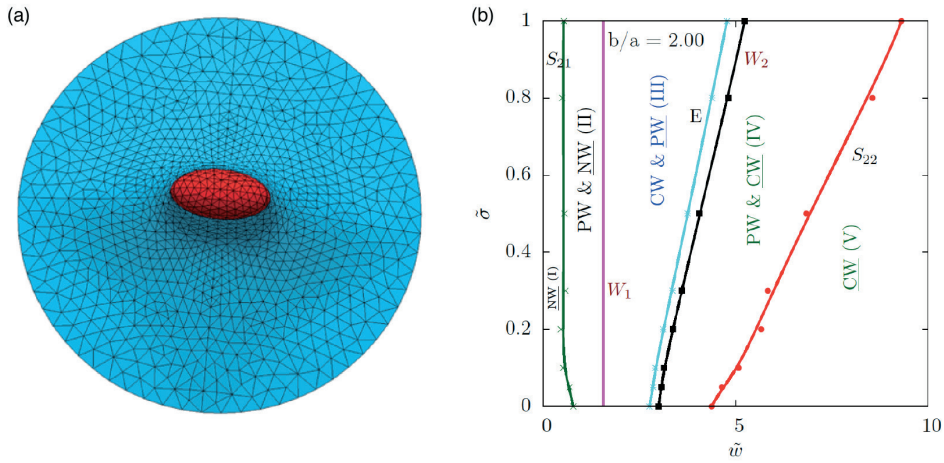


Fig. 8: *(a) Triangulated membrane with an attached ellipsoidal particle. (b) Wrapping states for a prolate ellipsoidal particles with aspect ratio* 2*, see Fig. 7 (b). Taken from Ref. [47].*

spherical particle forms a catenoid, which is a minimal surface with $H = 0$ at every point. Therefore, no bending energy costs are required for such a catenoidal deformation and a direct transition between a non-wrapped sphere and a completely wrapped sphere occurs for $\tilde{w} = 4$ also if the free membrane around the particle is taken into account. However, the transition to the completely-wrapped state estimated above is not accurate, due to the presence of a partially-wrapped state [46].

For both spherical particles wrapped by a membrane with bending rigidity and membrane tension and for non-spherical particles, the bending-energy increases nonlinearly with the wrapping fraction $A_{\rm ad}/A$ [39, 47, 46], see Fig. 7 (b). However, for homogeneous adhesion strength the adhesion energy gain increases linearly with the $A_{\rm ad}/A$. The total energy for wrapping a spherical particle is obtained by adding the deformation energy cost and the adhesion energy gain; the minimum of the energy indicates the wrapping state of the system. If the minimum of the energy is at $A_{\rm ad}/A = 0$, for a spherical particle for $\tilde{w} < 4$, the particle is not wrapped. For $4 < \tilde{w} \lesssim 6.12$, the particle is partially wrapped and for $6.12 \lesssim \tilde{w}$, the particle is completely wrapped. At an adhesion strength $\tilde{w} = 5.99$, the unwrapped state has the same energy as the completely-wrapped state, which can serve as estimate for complete wrapping. However, note that for this adhesion strength the partially-wrapped state is stable and separated from the completely-wrapped state by an energy barrier. This envelopment transition occurs for adhesion strength $\tilde{w} = 6.12$ with two minima of equal energy, which is analogous to a binodal in thermodynamics. The adhesion strengths where an energy barrier just vanishes and where wrapping or unwrapping occurs spontaneously define the wrapping and the unwrapping spinodal, respectively.

Fig. 8 (a) shows a partially-wrapped ellipsoidal particle attached to a planar membrane patch, the triangulation used for the numerical calculation of the energy is indicated. The membrane deformation energies have been calculated using Surface Evolver [38]. The phase diagram for a prolate ellipsoidal particle with aspect ratio 2 is shown in Fig. 8 (b), assuming that the particle is oriented with its long axis parallel to the membrane. While the binding transition W_1 is independent of the membrane tension, the adhesion strength for the envelopment transi-

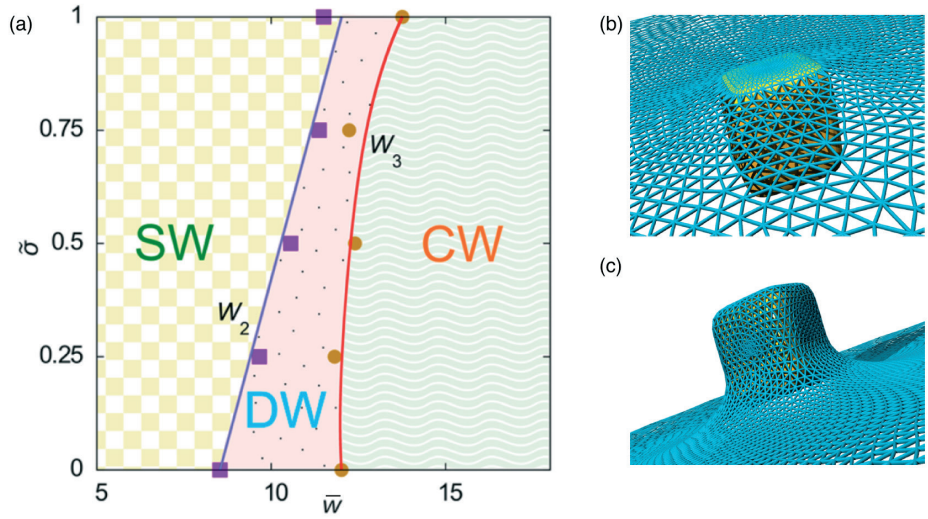
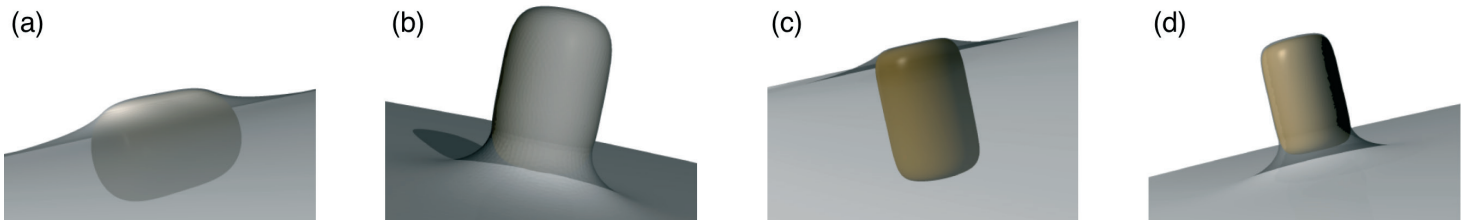


Fig. 9: *Wrapping of a cube-like nanoparticle. (a) Phase diagram for wrapping of Hauser's cube for membrane tension $\tilde{\sigma}$ and adhesion strength $\bar{w}$; the parameters are given in dimensionless form. We find a shallow-wrapped (SW), a deep-wrapped (DW), and a complete-wrapped (CW) state, separated by two discontinuous wrapping transitions, W_2 and W_3. (b), (c): Membrane deformation for wrapping of Hauser's cube. The network of edges and triangles describes the membrane shape and is used for the numerical calculation of the curvature energy. Membrane conformations are shown at fixed tension $\tilde{\sigma} = 0.50$ for two corresponding states at the W_2 phase boundary: (a) a shallow-wrapped state with approximately 10% of particle wrapped, and (b) a deep-wrapped state with a wrapping fraction of approximately 80%. Taken from Ref. [39].*

tion increases with the membrane tension [47, 46]. The spinodal for spontaneous wrapping is found at considerably higher energies than the envelopment transition, while the spinodal for spontaneous unwrapping is at lower adhesion strength than the binding transition. Spontaneous unwrapping therefore occurs directly to the non-wrapped state without a transition through a partially-wrapped state.

Unlike spherical and ellipsoidal nanoparticles, cube-like and and rod-like nanoparticles that are represented by superellipsoids (Fig. 2) have a point on the particle surface with vanishing curvature. The deformation energy cost at this point vanishes, therefore these particles bind immediately to a membrane already at infinitely small adhesion strength. The wrapping phase diagram for a cube-like particle shown in Fig. 9 (a) starts immediately with a shallow partially-wrapped state for small adhesion strengths. For increasing adhesion strength, first a deep-wrapped and finally a complete-wrapped state is found. The phase boundaries shift to increasing adhesion strength with increasing membrane tension. Fig. 9 (b) shows a shallow partially-wrapped state, Fig. 9 (c) a deep partially-wrapped state. For the transition between both partially-wrapped states and for the envelopment transition to the complete-wrapped state energy barriers need to be overcome, such that both transitions are discontinuos and associated with wrapping and unwrapping spinodals. For clarity, we only plot the transitions between the states and not the spinodals.

For elongated particles, not only the presence of partially-wrapped states and the distinction



particle shape	membrane	binding transition	shallow-wrapped state	deep-wrapping transition	deep-wrapped state	envelopment transition
spherical	κ	cont., for $w = 2\kappa/a^2$	-	-	-	$\equiv$ binding
spherical	κ and σ	cont., for $w = 2\kappa/a^2$	yes	-	-	discont.
ellipsoidal	κ, κ and σ	cont., indep. of σ	yes, submarine	discont., reorient.	yes, rocket	cont.
ellipsoidal (*)	κ, κ and σ	cont., indep. of σ	yes, submarine	-	-	discont.
cube-like	κ, κ and σ	at vanishing w	yes	discont.	yes	discont.
spherocylinder	κ, κ and σ	at vanishing w, rocket	yes, submarine	discont., reorient.	yes, rocket	discont.
rod-like	κ, κ and σ	at vanishing w, rocket	yes, submarine	discont., reorient.	yes, rocket	discont.
rod-like (*)	κ, κ and σ	at vanishing w, rocket	yes, submarine	-	-	discont.
rod-like (**)	κ, κ and σ	at vanishing w, rocket	yes, rocket	discont.	yes, rocket	discont.

Table 1: *Shape dependence of particle wrapping, based on Refs. [39, 47, 46]. Subfigures (a)-(d) show wrapping of nanorods in submarine and rocket orientation. The membrane can be characterised by bending rigidity only, 'κ', or by bending rigidity and membrane tension, 'κ and σ'; the binding transition can occur at finite or vanishing adhesion strength w; the particle can be in submarine or rocket orientation; transitions can be continuous (cont.) or discontinuous (discont.) and may involve reorientation (reorient.). The binding transition for ellipsoids is independent of the membrane tension and is given in Ref. [49]. (*) Fast wrapping at high adhesion strength, such that a bound ellipsoid cannot reorient to rocket orientation. (**) Rocket mode for supereggs with blunt tips and small aspect ratio (e. g. $n = 4$ and $b/a = 1.5$). Taken from Ref. [39].*

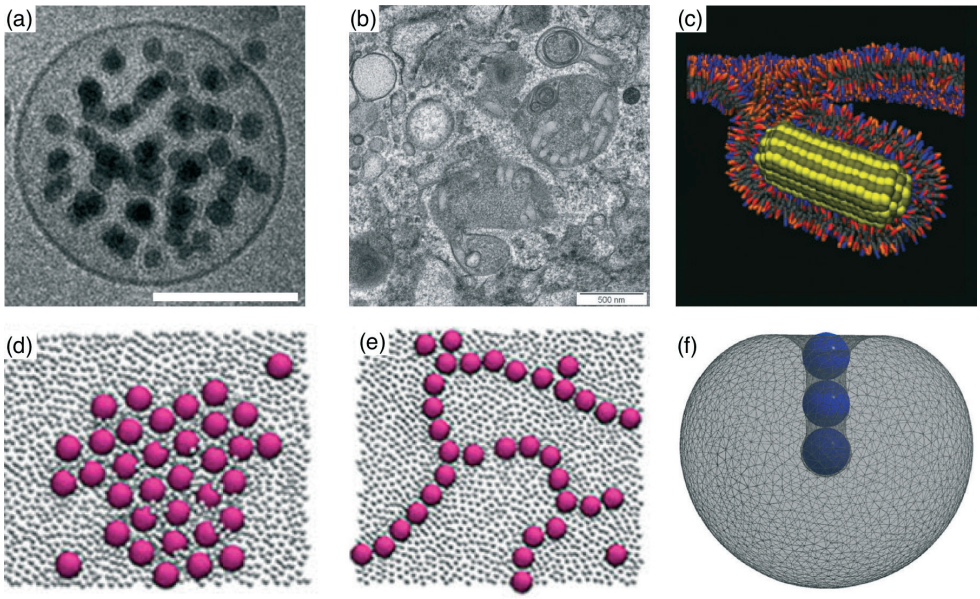


Fig. 10: *Particles at membranes. (a) Internalization of nanoparticles into a vesicle. The length of the scale bar corresponds to* 100 nm*. Taken from Ref. [21]. (b) Uptake of spherical and ellipsoidal polymeric nanoparticles by human mesenchymal stem cells. The length of the scale bar corresponds to* 500 nm*. Taken from Ref. [50]. (c) Wrapping of a nanoparticle by a lipid bilayer. Taken from Ref. [28]. (d) Nanoparticles adhered to a lipid bilayer membrane with (d) low and (e) medium bending rigidity. Taken from Ref. [51]. (f) Tubulation of nanoparticles adhered to a vesicle. Taken from Ref. [52].*

between a continuous transition without an energy barrier and a discontinuous transition with an energy barrier have to be taken into account, but also the orientation of the particle. The particle is in submarine orientation if its long axis is oriented parallel to the membrane, this orientation is preferentially found for shallow-wrapped states [39]. For high wrapping fraction, the particle usually reorients to rocket orientation with its long axis perpendicular to the membrane. An overview over wrapping of particles with different shape is provided in Tab. 1.

Up to now, we have only discussed the interaction of single nanoparticles with membranes. Figs. 10 (a)-(b) shows microscopy pictures for the interaction of many nanoparticles with a vesicle and for the incorporation of many nanoparticles into human mesenchymal stem cells, Figs. 10 (c) shows a simulation snapshot of a molecular simulation for wrapping an elongated nanoparticle. The Helfrich model and the phase diagrams presented earlier in this section provide a systematic approach to understand the interaction of the particles with lipid bilayer membranes. Furthermore, this mesoscopic approach allows to investigate the interaction of several nanoparticles with a membranes that mutually interact via membrane-mediated interactions.

Using computer simulations, different forms of aggregates are observed depending on the details of the model. For a membrane with a low bending rigidity, compact 2D aggregates have been observed, while for intermediate bending rigidities using the same model linear aggregates have been observed [51], see Fig. 10 (d) and (e); also tubulation is observed for the interaction of several nanoparticles with a membrane [52, 53], see Fig. 10 (f). There is only few quantita-

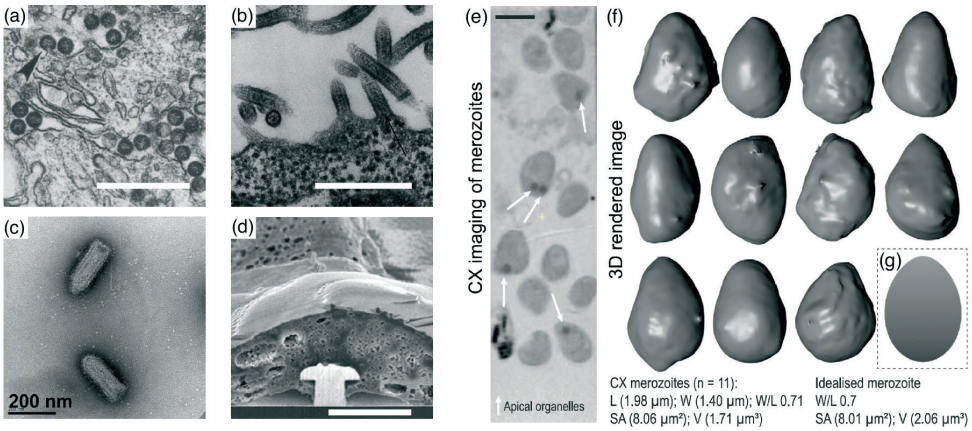


Fig. 11: *(a) Rift valley feaver viruses in primary rat hepatocytes. The length of the scale bar corresponds to* $600\,\mathrm{nm}$*. Taken from Ref. [8]. (b) Virus-like particles as seen in nucleocapsids of Ebola virions, about to bud. The length of the scale bar corresponds to* $500\,\mathrm{nm}$*. Taken from Ref. [9]. (c) Vesicular stomatitis virus (VSV). The length of the scale bar corresponds to* $200\,\mathrm{nm}$*. Taken from Ref. [10]. (d) A cell engulfing a 3D electrode for cell-chip coupling in bioelectronics. The length of the scale bar corresponds to* $1\,\mu\mathrm{m}$*. Taken from Ref. [58]. (e)-(g) Defining an idealized archetypal malaria parasite. (e) A section through cryo-x-ray imaged free P. falciparum merozoites cryopreserved in a capillary. Apical secretory organelles (specifically rhoptries) are visible as dense spots indicated by arrows. (f) Isosurface rendered merozoites from (e). (g) The idealized archetypal merozoite simulated as an asymmetrical egg-shaped rigid particle. Taken from Ref. [11].*

tive data available for several nanoparticles and a membrane, therefore the membrane-mediated interaction between nanoparticles is not well understood, but the related system of membrane-mediated interaction of curved inclusions has been studied in more detail. The interaction between two small inclusions in an almost planar membrane is always repulsive [54], many-particle interactions [55, 56] and large inclusions [57] can also lead to an attractive interaction.

4 Biological systems

In this section, we introduce biological systems related to the interaction of particles with membranes discussed in section 3. Viruses and parasites have a large variety of shapes and remind strongly of the nanoparticles shown in Fig. 1. Examples for viruses are shown in Fig. 11 (a)-(c), their sizes are in the range of tens to hundreds of nanometers. Fig. 11 (d) shows a cardiomyocyte adhered to a 3D gold electrode for recording of electrical currents of the cell; the cylindrical pillar of the electrode has a diameter of $0.3\,\mu m$ and a height of $0.3\,\mu m$. Fig. 11 (e)-(g) shows the malaria parasite with a typical length of $\approx 2\,\mu m$ and a width of $\approx 1.4\,\mu m$. The basic process for these systems is wrapping of a virus or a parasite by a lipid bilayer membrane. Compared with particle-membrane systems complexity is added for these systems, mainly because of the complexity of biological cells. The bilayer membrane for instance is usually a multi-component membrane [59], therefore studying the adhesion also has to take into account for specific and unspecific binding to discrete binding sites and for the entropy of such binding sites, e. g. of

receptors and ligands [60]. Furthermore, a cortical cytoskeleton might support the lipid bilayer that alters both the mechanical stability of the membrane [61, 34, 35] as well as the diffusion within the membrane [62]. A cytoskeleton that fills the entire volume of the cell may actively contribute to the wrapping process [63].

A particular example for a biological system that we will discuss in more detail is the invasion of the malaria parasite into a red blood cell [11]. Fig. 12 shows a schematics of the invasion of the malaria merozoite into the red blood cell. After an infected red blood cell bursts, the merozoites distribute in the blood plasma. They remain in a state where they are able to invade healthy red blood cells for about 10 minutes. Although merozoites contain molecular motors, the reorientation that precedes invasion may be explained solely by a gradient of adhesive force, for example due to an apical release of adhesive membrane proteins on the parasite. Only in the apical orientation the merozoite is able to successfully invade a red blood cell. Several processes have been reported or are hypothesized to accompany the invasion process, such as the formation of an electron-dense region upon junction formation, a disruption of the cytoskeleton of the red blood cell at the site of invasion, active pushing by motor forces, and the secretion of membrane from the inside of the parasite that may lead to a spontaneous curvature that supports wrapping.

To quantify the wrapping energy contributions to the invasion of the malaria parasite into a red blood cell, in addition to bending, tension, and adhesion contributions, a line-tension is used to model the electron-dense zone in the junction region.

$$\mathcal{E} = \mathcal{E}_{\text{bending}} + \mathcal{E}_{\text{membrane tension}} + \mathcal{E}_{\text{adhesion}} + \mathcal{E}_{\text{line tension}} \tag{7}$$

This model assumes that the spectrin cytoskeleton of the red blood cell is destroyed at the site where the merozoite enters the cell and intact at the junction zone between the merozoite and the red blood cell. Therefore a line tension at the junction zone may originate from stretching the spectrin cytoskeleton. While the bending and the tension energy are modeled as for wrapping of particles, the adhesion strength is chosen to be proportional to the curvature of the parasite surface.

$$\mathcal{E} = 2\kappa \int_{\text{RBC}} dS\,(H - c_0)^2 + \sigma \int_{\text{RBC}} dS - w \int_{\text{adhered}} dS\,H + \gamma \int_{\text{contact line}} dl\,. \tag{8}$$

The absolute values for the model parameters can be translated into dimensionless parameters using the radius of a sphere with the same surface area as the parasite, $a = 800\,\text{nm}$, as basic length scale of the system, and the membrane bending rigidity κ as energy scale. These dimensionless parameters indicated by a tilde are $\tilde{c}_0 = c_0 a^2 H_0$, $\tilde{\sigma} = \sigma a^2/(2\kappa)$, $\tilde{w} = w H_0 4\pi a^2/(2\kappa)$, and $\tilde{\gamma} = \gamma a/(2\kappa)$. The average mean curvature of the merozoite can be calculated as surface integral using the archetypal merozoite defined in the next section, $H_0 = \int_{\text{RBC}} dS\,H / \int_{\text{RBC}} dS = 2.5/a$. The spontaneous curvature can be used to construct an effective adhesion strength, $\tilde{w}_{\text{eff}} = \tilde{w} + \tilde{c}_0$, and an effective surface tension, $\tilde{\sigma}_{\text{eff}} = \tilde{\sigma} + \tilde{c}_0^2/(aH_0)^2$, such that the phases for different values of the spontaneous curvature can be extracted using effective parameter values.

Fig. 13 shows wrapping phase diagrams for the interaction of the merozoite with a red blood cell. Because of the egg-like shape of the parasite and the inhomogeneous adhesion strength, the phase diagrams are much richer than the phase diagrams for the interaction of particles with pure lipid-bilayer membranes. For example for a fixed line tension, the phase diagram for membrane tension and adhesion strength in Fig. 13 (a) shows two partially-wrapped states separated by an energy barrier; at a critical point this energy barrier vanishes. For low or effectively negative

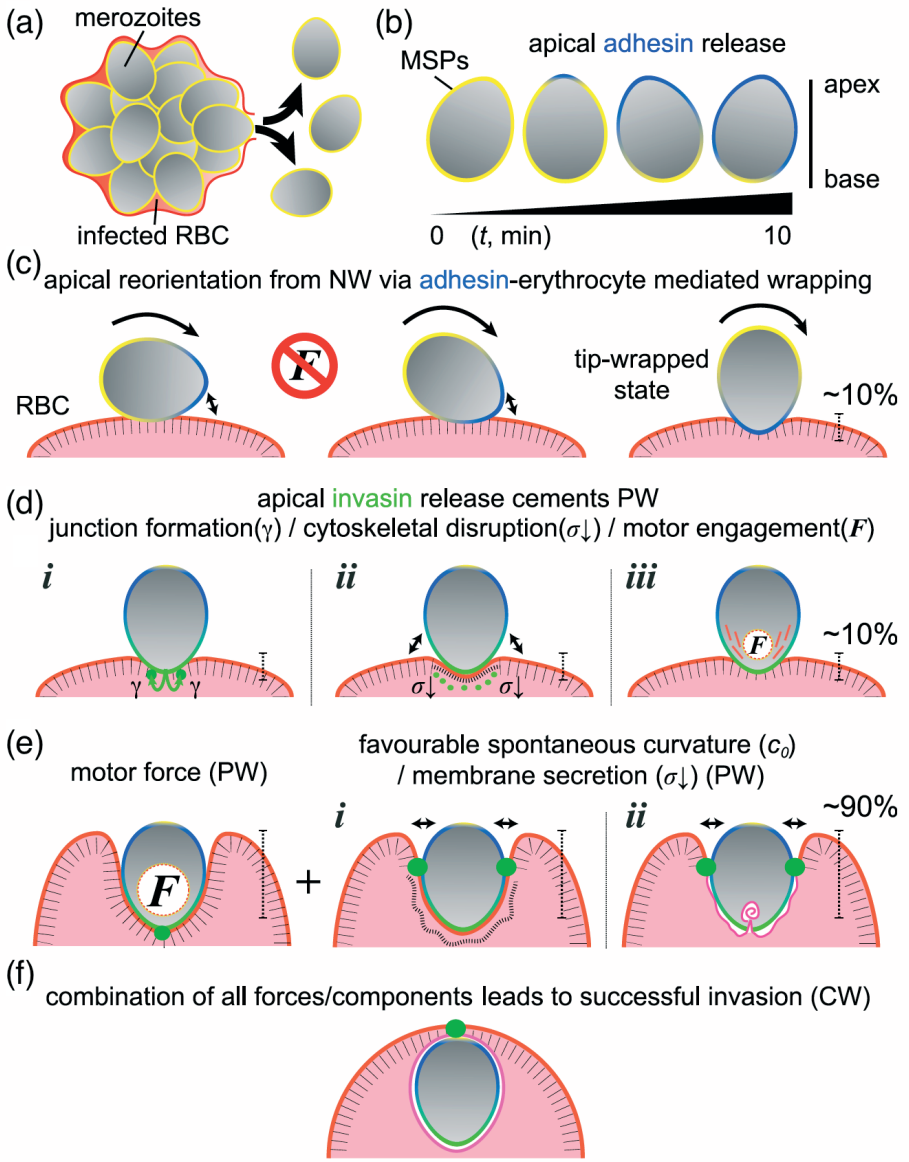


Fig. 12: *Schematic for biophysical interactions between the Plasmodium merozoite and the human erythrocyte. A complete biophysical model for merozoite invasion of the erythrocyte from release ((a) and (b)), to attachment and reorientation facilitating a stable, tip-wrapped state (c), to PW states ((d) and (e)), and full invasion/CW states (f). Taken from Ref. [11].*

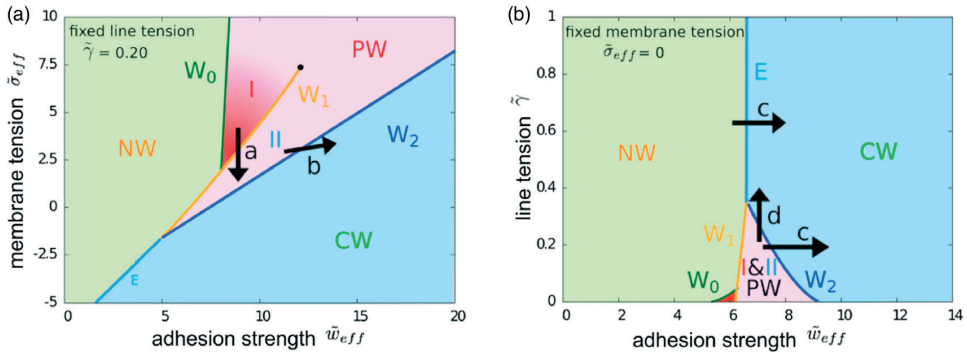


Fig. 13: *Wrapping phase diagrams for the interaction of the egg-shaped malaria parasite with a red blood cell. Wrapping phase diagram for fixed line tension or membrane tension. (a) Wrapping states of the system of a tip-first-oriented merozoite for fixed reduced line tension* $\tilde{\gamma} = 0.2$ *and several values of effective adhesion strength and effective membrane tension: non wrapped merozoite (NW), partially wrapped merozoite with small (PW I) and high wrapping fractions (PW II), and completely wrapped/fully invaded merozoite (CW). The transition* W_0 *is a continuous transition, whereas the transitions* W_1*,* W_2*, and* E *are associated with energy barriers. The transition* W_1 *ends at a critical point where the difference between PW I and PW II vanishes. (b) Wrapping states of the system of a tip-first-oriented merozoite for vanishing effective membrane tension,* $\tilde{\sigma}_{\text{eff}} = 0$*, and several values of effective adhesion strength and line tension. The notation is analogous to (a). The arrows indicate biological processes that are thought to accompany invasion: arrow a indicates the effect of unstructured membrane secreted by the merozoite, arrow b shows the effect of favorable spontaneous curvature, arrow c the effect of increased adhesion strength, and arrow d the effect of increased line tension. Taken from Ref. [11].*

membrane tension [64], partially-wrapped and completely-wrapped states are also found for adhesion strengths lower than those for the binding transition. The phase diagram for fixed membrane tension in Fig. 13 (b) shows that the partially-wrapped states vanish for high enough line tension. Arrows in the diagrams indicate the effect of active processes, such as secretion of unstructured membrane from the merozoite (arrow a), a change in the spontaneous curvature that favours wrapping (arrow b), an increased adhesion strength (arrow c), and an increased line tension (arrow d). The motor force that may be required to overcome energy barriers between the different states can also be quantified within the model depending on the path in the phase diagram. A typical value for the required motor force that is estimated from the energy barriers is $10 - 20\,\mathrm{pN}$.

5 Summary and outlook

A large variety of nanoparticles with various shapes, sizes, and surface properties can be produced and are used for applications. The interaction of micrometer-sized particles at interfaces has been studied both experimentally and theoretically and is well understood in particular for ellipsoidal particles. The particles deform the surrounding interface; for several particles the minimization of the total energy of the system and therefore the minimization of the interfa-

cial area leads to long-ranged capillary forces. Due to the large interaction energies, the final, self-assembled network structure that is observed in experiments can be metastable.
Less than for the capillary interaction is known for the interaction of (nano-)particles with membranes, although theoretically the interaction of a single nanoparticle with a membrane has been well studied. Spherical nanoparticles at fluid membranes without membrane tension are either unwrapped or–beyond a threshold adhesion strength–completely wrapped. For membranes with tension, in addition partially-wrapped states are found. Partially-wrapped states also occur for non-spherical nanoparticles, such as ellipsoidal, cube-like, and rod-like nanoparticles. The transitions between the partially-wrapped states can be either continous and discontinous, for elongated particles also the orientation of the particle matters for the wrapping process. For the interaction of many nanoparticles with a membrane, two-dimensional compact aggregates and linear aggregates on the membrane, as well as tubulation has been observed. However, a complete understanding of cooperative interaction of many nanoparticles is still lacking.
Biological systems are complex: in addition to passive wrapping of a particle with homogenous lipid-bilayer membrane, discrete adhesion sites such as receptors, as well as active cytoskeletal processes have to be taken into account. Effective cytoskeletal models have been used to explain particle uptake by biological cells, but there is need for a detailed understanding of the interplay of particles, membranes, receptors, and cytoskeleton.
Acknowledgements. These lecture notes are based on my work together with Sabyasachi Dasgupta and Gerhard Gompper, in particular on Refs. [47], [39], and [11].

Appendices

A Differential geometry and the Helfrich Hamiltonian

Differential geometry

The basic technique used on this level of description is differential geometry [65, 66, 67, 68]. Based on Ref. [65], we will first discuss a curve in a plane and then a surface embedded in the three-dimensional space.

A curve in $\mathbb{R}^2$

For a curve in a plane that is defined by the real function $\boldsymbol{\alpha}(t) = (a_1(t), a_2(t))$ on the open interval $(a, b) \in \mathbb{R}$, the velocity of the curve is given by $\boldsymbol{\alpha}'(t) = (a_1'(t), a_2'(t))$. We can define the local curvature,

$$c(t) = \frac{\boldsymbol{\alpha}''(t) J \boldsymbol{\alpha}'(t)}{||\boldsymbol{\alpha}'(t)||^3}\,, \tag{9}$$

where $J(p_1, p_2)$ is the complex structure on $\mathbb{R}^2$, $J(p_1, p_2) = (-p_2, p_1)$, and $||\boldsymbol{p}|| = \sqrt{\boldsymbol{p} \cdot \boldsymbol{p}}$. For $\boldsymbol{\alpha}(t) = (x(t), y(t))$,

$$c(t) = \frac{x'(t)y''(t) - x''(t)y'(t)}{x'^2(t) + y'^2(t)}\,. \tag{10}$$

The curvature reflects the infinitesimal change of direction of the tangent vector along the curve and thus also of the normal vector, $d\boldsymbol{n} = c\, d\boldsymbol{x}$, see Fig. 14. The sign of the curvature depends on the direction of the normal vector and in Eq. (10), $c(t) = -c(-t)$. The absolute value of the curvature is the inverse of the radius of the osculating circle that is locally fit to the curve

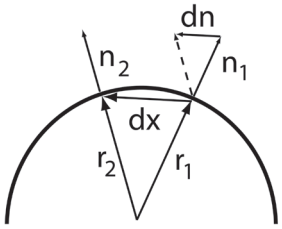


Fig. 14: *Infinitesimal changes of the normal vector are given by the curvature c.*

and has the center $\boldsymbol{\alpha}_0(t) = \boldsymbol{\alpha}(t) + (1/c(t))(J\boldsymbol{\alpha}'(t)/||\boldsymbol{\alpha}'(t)||)$. For a circle $x^2 + y^2 = a_\mathrm{c}^2$ with $\boldsymbol{\alpha}_\mathbf{c}(t) = (a_\mathrm{c}\cos t, a_\mathrm{c}\sin t)$ for $0 \le t \le 2\pi$, we obtain $c = 1/a_\mathrm{c}$ and $\boldsymbol{\alpha}_0(t) = (0,0)$.
For $\boldsymbol{\alpha}_\mathbf{s}(t) = (t, y(t))$, we find $c(t) = y''(t)/(1+y'^2(t))$. In the small-gradient approximation, if $\boldsymbol{\alpha}_\mathbf{s}(t)$ describes an almost straight and horizontal line, $1/(1+y'^2(t)) = 1+O[y'^2(t)]$ and $c(t) \approx y''(t)$. For example, the curvature of a sine function with amplitude a_s, $\alpha_\mathrm{s}(\mathrm{t}) = (\mathrm{t}, \mathrm{a_s}\sin(\mathrm{t}))$, vanishes at the straight parts of the curve where $y(t) = y''(t) = 0$ and its maxima and minima correspond to the maxima and the minima of $\sin t$: $c(t) = -a_\mathrm{s}\sin t/(1 + a_\mathrm{s}^2\cos^2 t)$. If the amplitude of the sine function is small, $c(t) \approx -a_\mathrm{s}\sin(t)$.

A surface in $\mathbb{R}^3$

In order to describe change of the normal vector for a small displacement on the surface, a scalar is not any more sufficient. Besides for the special case of a sphere, the cuts of planes that contain the surface normal vector with the surface itself are curves that have osculating circles of different radii. In general, the infinitesimal change of the normal vector, $d\boldsymbol{n}$, is not parallel to the deplacement on the surface, $d\boldsymbol{x}$. We need to go beyond the level of simple algebraic operations and the curvature is replaced by the curvature tensor $\mathfrak{C}$,

$$d\boldsymbol{n} = \mathfrak{C}\, d\boldsymbol{x}\,, \tag{11}$$

which is a 2×2 matrix. To develop a suitable formalism, we first define a tangential vector of the surface $f(u_1, u_2, u_3)$ in the point $\boldsymbol{p}$ along a line given by $\boldsymbol{p} + t\boldsymbol{v} = (p_1, p_2, p_3) + t\,(v_1, v_2, v_3)$,

$$\boldsymbol{v}_\mathrm{p}[f] = \frac{d}{dt}\,(f(\boldsymbol{p} + t\boldsymbol{v}))|_{t=0} = \left(\sum_{j=1}^{3} v_j \frac{\partial f}{\partial u_j}\,(\boldsymbol{p})\right)\bigg|_{t=0}\,. \tag{12}$$

The tangent space for the surface that is spanned by the $\boldsymbol{v}_\mathrm{p}$ is isomorphic to the surface itself. To deal with a function $F(g_1(\boldsymbol{p}), g_2(\boldsymbol{p}), g_3(\boldsymbol{p}))$, we can use the chain rule for differentiation,

$$\boldsymbol{v}_p[f] = \sum_{i=1}^{3}\sum_{j=1,2} v_i \frac{\partial F}{\partial g_j}(\boldsymbol{g}(\boldsymbol{p}))\,\frac{\partial g_j}{\partial u_i}(\boldsymbol{p}) = \sum_{j=1,2}\frac{\partial F}{\partial g_j}(\boldsymbol{g}(\boldsymbol{p}))\,\boldsymbol{v}_p[g_j]\,. \tag{13}$$

The terms $\partial g_j/\partial u_i$ are often given in form of the Jacobian matrix, e. g.

$$\mathfrak{J}(g_1, g_2, g_3)(u_1, u_2) = \begin{pmatrix} \partial g_1/\partial u_1 & \partial g_2/\partial u_1 & \partial g_3/\partial u_1 \\ \partial g_1/\partial u_2 & \partial g_2/\partial u_2 & \partial g_3/\partial u_3 \end{pmatrix}\,. \tag{14}$$

Very often, a height field is used to describe a membrane conformation (Monge gauge or Monge representation of the surface),

$$\boldsymbol{x}_{\mathrm{M}}(u,v) = (u, v, h(u,v))\,, \tag{15}$$

which is appropriate for every system if only a small-enough membrane patch is chosen.
Let us now define the curvature of a surface, which can be quantified by the change of the orientation of the normal vector. A central element is the shape operator $S(\,)$, which acts on a tangential vector $\boldsymbol{v}_p$ and gives the negative derivative of the normal vector along $\boldsymbol{v}_p$. The normal curvature in the direction of a tangential vector $\boldsymbol{v}_p$ is then defined as

$$\boldsymbol{k}(\boldsymbol{v}_p) = \frac{S(\boldsymbol{v}_p)\cdot\boldsymbol{v}_p}{||\boldsymbol{v}_p||^2}\,. \tag{16}$$

We define the second fundamental form,

$$\begin{aligned} e &= S(\boldsymbol{x}_u)\cdot\boldsymbol{x}_u = -\boldsymbol{n}_u\cdot\boldsymbol{x}_u = \boldsymbol{n}\cdot\boldsymbol{x}_{uu} && (17)\\ f &= S(\boldsymbol{x}_v)\cdot\boldsymbol{x}_u = -\boldsymbol{n}_v\cdot\boldsymbol{x}_u = \boldsymbol{n}\cdot\boldsymbol{x}_{uv} = \boldsymbol{n}\cdot\boldsymbol{x}_{vu} = -\boldsymbol{n}_u\cdot\boldsymbol{x}_v = S(\boldsymbol{x}_u)\cdot\boldsymbol{x}_v && (18)\\ g &= S(\boldsymbol{x}_v)\cdot\boldsymbol{x}_v = -\boldsymbol{n}_v\cdot\boldsymbol{x}_v = \boldsymbol{n}\cdot\boldsymbol{x}_{vv}\,, && (19) \end{aligned}$$

where the orthogonality of the tangent vectors and the normal vector has been used, e. g.

$$0 = \frac{\partial}{\partial v}\left(\boldsymbol{n}\cdot\boldsymbol{x}_u\right) = \boldsymbol{n}_v\cdot\boldsymbol{x}_u + \boldsymbol{n}\cdot\boldsymbol{x}_{uv} \tag{20}$$

and analogous expressions hold for other combinations of the derivatives.
We can express $S(\,)$ through the first and the second fundamental form: In a regular coordinate system, the basis vectors $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ are linearly independent and we can use the ansatz

$$\begin{aligned} -S(\boldsymbol{x}_u) &= \boldsymbol{n}_u = c_{11}\boldsymbol{x}_u + c_{12}\boldsymbol{x}_v && (21)\\ -S(\boldsymbol{x}_v) &= \boldsymbol{n}_v = c_{21}\boldsymbol{x}_u + c_{22}\boldsymbol{x}_v\,. \end{aligned}$$

We project Eqs. (21) and (22) to the basis vectors $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ to obtain a set of four equations,

$$\begin{aligned} -e &= c_{11}E + c_{12}F && (22)\\ -f &= c_{11}F + c_{12}G\\ -f &= c_{21}E + c_{22}F\\ -g &= c_{21}F + c_{22}G\,. \end{aligned}$$

Solving the matrix equation for the curvature tensor,

$$\mathfrak{C} = \begin{pmatrix} c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix} = -\begin{pmatrix} e & f\\ f & g\end{pmatrix}\begin{pmatrix} E & F\\ F & G\end{pmatrix}^{-1} = -\frac{1}{EG-F^2}\begin{pmatrix} e & f\\ f & g\end{pmatrix}\begin{pmatrix} G & -F\\ -F & E\end{pmatrix}\,, \tag{23}$$

we obtain

$$\begin{aligned} -S(\boldsymbol{x}_u) &= \boldsymbol{n}_u = \frac{fF-eG}{EG-F^2}\boldsymbol{x}_u + \frac{eF-fE}{EG-F^2}\boldsymbol{x}_v && (24)\\ -S(\boldsymbol{x}_u) &= \boldsymbol{n}_v = \frac{gF-fG}{EG-F^2}\boldsymbol{x}_u + \frac{fF-gE}{EG-F^2}\boldsymbol{x}_v\,. \end{aligned}$$

The normal curvature in the direction $\boldsymbol{v}_p = a\boldsymbol{x}_u + b\boldsymbol{x}_v$ can be written as

$$\boldsymbol{k}(\boldsymbol{v}_p) = \frac{ea^2 + 2fab + gb^2}{Ea^2 + 2Fab + Gb^2}\,. \tag{25}$$

The curvature tensor (Eq. (23)) / the shape operator $S(\,)$ is symmetric (self-adjoint, Eq. (18)), therefore its eigenvalues λ_i are real numbers. If $\hat{\boldsymbol{e}}_1$ and $\hat{\boldsymbol{e}}_1$ are the (orthonormal) eigenvectors with $S(\hat{\boldsymbol{e}}_1) = \lambda_1\hat{\boldsymbol{e}}_1$ and $S(\hat{\boldsymbol{e}}_2) = \lambda_2\hat{\boldsymbol{e}}_2$, using the ansatz

$$\boldsymbol{u}_i(\theta) = \cos(\theta)\hat{\boldsymbol{e}}_1 + \sin(\theta)\hat{\boldsymbol{e}}_2\,, \tag{26}$$

we obtain

$$S(\boldsymbol{u}_i) = \cos(\theta)S(\hat{\boldsymbol{e}}_1) + \sin(\theta)S(\hat{\boldsymbol{e}}_2) = \lambda_i(\theta)(\cos(\theta)\hat{\boldsymbol{e}}_1 + \sin(\theta)\hat{\boldsymbol{e}}_2)\,, \tag{27}$$

$$\lambda_i(\theta) = \cos^2(\theta)S(\hat{\boldsymbol{e}}_1)\cdot\hat{\boldsymbol{e}}_1 + 2\sin(\theta)\cos(\theta)S(\hat{\boldsymbol{e}}_2)\cdot\hat{\boldsymbol{e}}_1 + \sin^2(\theta)S(\hat{\boldsymbol{e}}_2)\cdot\hat{\boldsymbol{e}}_2\,, \tag{28}$$

and $\lambda_1 \geq \lambda(\theta) \geq \lambda_2$. The eigenvalues correspond to those values θ_i for which the derivative

$$\frac{d\lambda}{d\theta}(\theta) = 2(S(\hat{\boldsymbol{e}}_2)\cdot\hat{\boldsymbol{e}}_2 - S(\hat{\boldsymbol{e}}_1)\cdot\hat{\boldsymbol{e}}_1)\sin(\theta)\cos(\theta) + 2S(\hat{\boldsymbol{e}}_1)\cdot\hat{\boldsymbol{e}}_2(\cos^2(\theta) - \sin^2(\theta)) \tag{29}$$

vanishes. Because $2S(\hat{\boldsymbol{e}}_1)\cdot\hat{\boldsymbol{e}}_2 = 0$ and, in general, $S(\hat{\boldsymbol{e}}_2)\cdot\hat{\boldsymbol{e}}_2 - S(\hat{\boldsymbol{e}}_1)\cdot\hat{\boldsymbol{e}}_1 \neq 0$, the eigenvalues require $\sin\theta\cos\theta = 0$ and $\theta_1 = 0$ or $\theta_2 = \pi/2$. The eigenvalues are the maximal and the minimal normal curvature (Eq. (16)) at each point of the surface, $c_1 = \lambda(\theta_1) = S(\hat{\boldsymbol{e}}_1)\cdot\hat{\boldsymbol{e}}_1$ and $c_2 = \lambda(\theta_2) = S(\hat{\boldsymbol{e}}_2)\cdot\hat{\boldsymbol{e}}_2$, and are called principal curvatures.

Mean and Gaussian curvature

From linear algebra, we know that for the determinant of a matrix a,

$$\det(U^{-1}\,a\,U) = \det(U)^{-1}\det(a)\det(U) = \det(a)\,, \tag{30}$$

and for the trace, i. e. the sum of the diagonal elements,

$$\mathrm{tr}(U^{-1}aU) = \mathrm{tr}(aUU^{-1}) = \mathrm{tr}(a)\,. \tag{31}$$

Therefore, the determinant and the trace of a are invariant with respect to the choice of the coordinate system. In mesoscopic membrane models, trace and the determinant of the curvature tensor at the point $\boldsymbol{p}$, mean curvature H and Gaussian curvature K,

$$H(\boldsymbol{p}) = \frac{1}{2}\mathrm{tr}(\mathfrak{C}) = \frac{1}{2}\mathrm{tr}(S(\boldsymbol{p})) = \frac{eG - 2fF + gE}{2(EG - F^2)} = \frac{c_1 + c_2}{2} \tag{32}$$

$$K(\boldsymbol{p}) = \det(\mathfrak{C}) = \det(S(\boldsymbol{p})) = \frac{eg - f^2}{EG - F^2} = c_1c_2\,, \tag{33}$$

are used to characterize the membrane shape. Note that K is independent of the choice of the normal vector of the surface, while the sign of H (and of the spontaneous curvature c_0) depends on the orientation of the surface and is a matter of definition. The principal curvatures are solutions of the quadratic equation $c^2 - 2Hc + K = 0$ and can be expressed by H and K,

$$c_{1,2} = H \pm \sqrt{H^2 - K}\,. \tag{34}$$

For some special surfaces, the principal curvatures and H and K are immediately obvious:

- For a sphere of radius R, $c_1 = c_2 = 1/R$, $H = 1/R$, $K = 1/R^2$.
- For a cylinder of radius R, $c_1 = 1/R$, $c_2 = 0$, $H = 1/(2R)$, $K = 0$.
- For a plane, $c_1 = c_2 = 0$, $H = K = 0$.

Usually, mean and Gaussian curvature are calculated using E, F, G, e, f, and g. For example, for a membrane in the Monge representation, $\boldsymbol{x}(u, v) = (u, v, h(u, v))$,

$$\boldsymbol{x}_u = (1, 0, \partial h/\partial u) \qquad \boldsymbol{x}_v = (0, 1, \partial h/\partial v) \qquad \boldsymbol{n} = \frac{\boldsymbol{x}_u \times \boldsymbol{x}_v}{||\boldsymbol{x}_u \times \boldsymbol{x}_v||} = (-h_u, -h_v, 1) \tag{35}$$

$$E = \boldsymbol{x}_u \cdot \boldsymbol{x}_u = 1 + h_u^2 \qquad F = \boldsymbol{x}_u \cdot \boldsymbol{x}_v = h_u h_v \qquad G = \boldsymbol{x}_v \cdot \boldsymbol{x}_v = 1 + h_v^2$$
$$e = \boldsymbol{n} \cdot \boldsymbol{x}_{uu} = h_{uu} \qquad f = \boldsymbol{n} \cdot \boldsymbol{x}_{uv} = h_{uv} \qquad g = \boldsymbol{n} \cdot \boldsymbol{x}_{vv} = h_{vv}$$

$$\begin{aligned} H &= \frac{(1 + h_u^2)h_{vv} + (1 + h_v)^2 h_{uu} - 2h_u h_v h_{uv}}{2\left(1 + h_u^2 + h_v^2\right)^{3/2}} \\ K &= \frac{h_{uu}h_{vv} - h_{uv}^2}{(1 + h_u^2 + h_v^2)^2}\,. \end{aligned}$$

Often x and y are used instead of u and v to denote the cartesian coordinates and for many applications, a small-gradient expansion for almost planar membranes ($h_x \ll 1$, $h_y \ll 1$) is sufficient:

$$H \approx \frac{1}{2}\left(h_{xx} + h_{yy}\right) = \frac{1}{2}\nabla^2 h \tag{36}$$
$$K \approx h_{xx}h_{yy} - h_{xy}^2\,. \tag{37}$$

Helfrich Hamiltonian

The shape of a membrane is characterized by the mean curvature, $H = (c_1 + c_2)/2$, and the Gaussian curvature, $K = c_1 c_2$. The minimal and the maximal curvature at each point of the membrane, c_1 and c_2, are called the principal curvatures. Using H and K, we can construct the energy density for a curved membrane,

$$\frac{\mathcal{E}}{A} = 2\kappa(H - c_0)^2 + \bar{\kappa}K \tag{38}$$

at each point of the membrane, the integral over the entire membrane area gives the system's energy. The three curvature elastic constants κ, c_0, and $\bar{\kappa}$ are the bending rigidity, the spontaneous curvature, and the Gaussian saddle splay modulus respectively. They link the mathematical surface to a physical membrane. This Hamiltonian is often referred to as "Helfrich Hamiltonian" or "Canham-Helfrich Hamiltonian" [33, 69].
The curvature-elastic constants, the material parameters for the membrane, are determined from experimental data. The bending rigidity can for example be extracted from fluctuation measurements [70, 71, 72] and a homogeneous spontaneous curvature corresponds to the vesicle shape [73, 74]. A homogeneous Gaussian saddle splay modulus of a membrane affects the system's energy only if the topology of the object changes, therefore not many direct measurements are available. Using stability arguments, it can be shown that $-2\kappa \leq \bar{\kappa} \leq 0$ [75].

B Triangulated membranes

Triangulation is a powerful method to calculate integrals over surfaces for arbitrary shapes. For many of the calculations presented in this manuscript, the freely available program package Surface Evolver [38] has been used that minimizes the energy of a triangulated surface for various energy functionals. While the energy of interfaces and the membrane tension contribution are proportional to the interfacial or the membrane area, respectively, an adhesion energy only applies to those triangles that are in contact with a particle. Also the deformation energy of a membrane can be calculated on a triangular network with high accuracy [49, 37]. For the discretization of the bending energy given by the Helfrich Hamiltonian, for example the Surface Evolver algorithm "star_perp_sq_mean_curvature" can be used. In this discretization, the bending energy is given by

$$\mathcal{E}_{\mathrm{bend}} = 2\kappa \sum_{v=1}^{n} a_v h_v^2 \tag{39}$$

with the local mean curvature at each vertex v,

$$h_v = \frac{1}{2} \frac{\boldsymbol{\nabla} a_v \cdot \boldsymbol{\nabla} \mathcal{V}_v}{\boldsymbol{\nabla} \mathcal{V}_v \cdot \boldsymbol{\nabla} \mathcal{V}_v} \,. \tag{40}$$

The area associated with each vertex is a_v and the volume is

$$\mathcal{V}_v = \frac{1}{6} \boldsymbol{v} \cdot \left[\boldsymbol{v}^1 \times \boldsymbol{v}^2 + \boldsymbol{v}^2 \times \boldsymbol{v}^3 + ... + \boldsymbol{v}^n \times \boldsymbol{v}^1 \right] , \tag{41}$$

such that

$$\boldsymbol{\nabla} \mathcal{V}_v = \frac{1}{6} \left[\boldsymbol{v}^1 \times \boldsymbol{v}^2 + \boldsymbol{v}^2 \times \boldsymbol{v}^3 + ... + \boldsymbol{v}^n \times \boldsymbol{v}^1 \right] , \tag{42}$$

where the postition vector for the vertex v is $\boldsymbol{v}$, while $\boldsymbol{v}^1$, $\boldsymbol{v}^2$, $\boldsymbol{v}^3$, ..., $\boldsymbol{v}^n$ are the position vectors for the neighboring vertices.

References

[1] K. S. Soppimath, T. M. Aminabhavi, A. R. Kulkarni, and W. E. Rudzinski, J. Controlled Release **70**(1), 1 (2001).

[2] R. H. Müller, K. Mäder, and S. Gohla, Eur. J. Pharm. Biopharm. **50**(1), 161 (2000).

[3] R. Hergt, S. Dutz, R. Müller, and M. Zeisberger, J. Phys. Condens. Matter **18**(38), S2919 (2006).

[4] V. K. Sharma, R. A. Yngard, and Y. Lin, Adv. Coll. Int. Sci. **145**(1), 83 (2009).

[5] L. Elias, F. Fenouillot, J.-C. Majesté, and P. Cassagnau, Polymer **48**(20), 6029 (2007).

[6] V. Tohver, A. Chan, O. Sakurada, and J. A. Lewis, Langmuir **17**(26), 8414 (2001).

[7] J. Ubbink, A. Burbidge, and R. Mezzenga, Soft Matter **4**(8), 1569 (2008).

[8] G. W. Anderson Jr and J. F. Smith, Virology **161**(1), 91 (1987).

[9] T. Noda, H. Ebihara, Y. Muramoto, K. Fujii, A. Takada, H. Sagara, J. H. Kim, H. Kida, H. Feldmann, and Y. Kawaoka, PLoS Pathog. **2**(9), 9 (2006).

[10] D. K. Cureton, R. H. Massol, S. P. Whelan, and T. Kirchhausen, PLoS Pathog. **6**(9), e1001127 (2010).

[11] S. Dasgupta, T. Auth, N. S. Gov, T. J. Satchwell, E. Hanssen, E. S. Zuccala, D. T. Riglar, A. M. Toye, T. Betz, J. Baum, *et al.*, Biophys. J. **107**(1), 43 (2014).

[12] M. Wortis and R. Mukhopadhyay, Proceedings of the National Academy of Sciences **99**(26), 16766 (2002).

[13] J. K. Dhont, in *Studies in interface science*, edited by D. Möbius and R. Miller (Elsevier, 1996).

[14] H. Lehle, E. Noruzifar, and M. Oettel, Eur. Phys. J. E **26**(1), 151 (2008).

[15] J. Loudet and B. Pouligny, Eur. Phys. J. E **34(8)**(8), 1 (2011), ISSN 1292-8941.

[16] B. Madivala, S. Vandebril, J. Fransaer, and J. Vermant, Soft Matter **5**, 1717 (2009).

[17] M. Cavallaro, L. Botto, E. P. Lewandowski, M. Wang, and K. J. Stebe, Proc. Natl. Acad. Sci. U. S. A. **108**(52), 20923 (2011).

[18] D. Ershov, J. Sprakel, J. Appel, M. A. C. Stuart, and J. van der Gucht, Proc. Natl. Acad. Sci. U. S. A. **110**(23), 9220 (2013).

[19] D. Lee and D. A. Weitz, Small **5**(17), 1932 (2009).

[20] S. U. Pickering, J. Chem. Soc., Trans. **91**, 2001 (1907).

[21] R. Michel, E. Kesselman, T. Plostica, D. Danino, and M. Gradzielski, Angew. Chem. Int. Ed. **53**(46), 12441 (2014).

[22] K. Jaskiewicz, A. Larsen, I. Lieberwirth, K. Koynov, W. Meier, G. Fytas, A. Kroeger, and K. Landfester, Angew. Chem. **124**(19), 4691 (2012).

[23] K. Jaskiewicz, A. Larsen, D. Schaeffel, K. Koynov, I. Lieberwirth, G. Fytas, K. Landfester, and A. Kroeger, ACS Nano **6**(8), 7254 (2012).

[24] F. Tian, X. Zhang, and W. Dong, Phys. Rev. E **90**(5), 052701 (2014).

[25] J. Barnoud, G. Rossi, and L. Monticelli, Phys. Rev. Lett. **112**(6), 068102 (2014).

[26] S. Pogodin, M. Werner, J.-U. Sommer, and V. A. Baulin, ACS Nano **6**(12), 10555 (2012).

[27] K. Yang and Y.-Q. Ma, Nat. Nanotech. **5**(8), 579 (2010).

[28] R. Vácha, F. J. Martinez Veracoechea, and D. Frenkel, Nano Lett. **11**(12), 5391 (2011).

[29] B. J. Reynwar, G. Illya, V. A. Harmandaris, M. M. Müller, K. Kremer, and M. Deserno, Nature **447**(7143), 461 (2007).

[30] J. A. Champion, Y. K. Katare, and S. Mitragotri, Proc. Natl. Acad. Sci. U. S. A. **104**(29), 11901 (2007).

[31] J.-G. Park, J. D. Forster, and E. R. Dufresne, J. Am. Chem. Soc. **132**(17), 5960 (2010).

[32] E. C. Cho, Q. Zhang, and Y. Xia, Nat. Nanotech. **6**(6), 385 (2011).

[33] W. Helfrich, Z. Naturforsch. C **28**(11), 693 (1973).

[34] T. Auth, S. A. Safran, and N. S. Gov, Phys. Rev. E **76**, 051910 (2007).

[35] J. Hansen, R. Skalak, S. Chien, and A. Hoger, Biophys. J. **72**(5), 2369 (1997).

[36] A.-S. Smith, E. Sackmann, and U. Seifert, Europhys. Lett.) **64**(2), 281 (2003).

[37] G. Gompper and D. M. Kroll, in *Statistical Mechanics of Membranes and Surfaces*, edited by D. R. Nelson, T. Piran, and S. Weinberg (World Scientific, Singapore, 2004).

[38] K. A. Brakke, Exp. Math. **1**(2), 141 (1992).

[39] S. Dasgupta, T. Auth, and G. Gompper, Nano Lett. **14**(2), 687 (2014).

[40] S. Dasgupta, M. Katava, M. Faraj, T. Auth, and G. Gompper, Langmuir **30**(40), 11873 (2014).

[41] J. C. Loudet and B. Pouligny, Europhys. Lett. **85**(2), 28003 (2009).

[42] M. Oettel and S. Dietrich, Langmuir **24**(4), 1425 (2008).

[43] M. Oettel, A. Dominguez, and S. Dietrich, Phys. Rev. E **71**, 051401 (2005).

[44] M. Nikolaides, A. Bausch, M. Hsu, A. Dinsmore, M. P. Brenner, C. Gay, and D. A. Weitz, Nature **420**(6913), 299 (2002).

[45] E. P. Lewandowski, M. Cavallaro, L. Botto, J. C. Bernate, V. Garbin, and K. J. Stebe, Langmuir **26**(19), 15142 (2010).

[46] M. Deserno, Phys. Rev. E. **69**, 031903 (2003).

[47] S. Dasgupta, T. Auth, and G. Gompper, Soft Matter **9**, 5473 (2013).

[48] M. Deserno, J. Phys.: Condens. Matter **16**, S2061 (2004).

[49] S. Dasgupta, T. Auth, and G. Gompper, Soft Matter **9**, 5473 (2013).

[50] L. Florez, C. Herrmann, J. M. Cramer, C. P. Hauser, K. Koynov, K. Landfester, D. Crespy, and V. Mailänder, Small **8**(14), 2222 (2012).

[51] A. Šarić and A. Cacciuto, Phys. Rev. Lett. **108**(11), 118101 (2012).

[52] A. H. Bahrami, R. Lipowsky, and T. R. Weikl, Phys. Rev. Lett. **109**(18), 188102 (2012).

[53] A. Šarić and A. Cacciuto, Physical review letters **109**(18), 188101 (2012).

[54] M. Goulian, R. Bruinsma, and P. Pincus, Europhys. Lett. **22**(2), 145 (1993).

[55] T. Auth and G. Gompper, Phys. Rev. E **80**(3), 031901 (2009).

[56] K. Kim, J. Neu, and G. Oster, Biophys. J. **75**(5), 2274 (1998).

[57] B. J. Reynwar and M. Deserno, Soft Matter **7**(18), 8567 (2011).

[58] F. Santoro, S. Dasgupta, J. Schnitker, T. Auth, E. Neumann, G. Panaitov, G. Gompper, and A. Offenhäusser, ACS nano **8**(7), 6713 (2014).

[59] E. Sackmann, in *Structure and Dynamics of Membranes*, edited by R. Lipowsky and E. Sackmann (Elsevier, Amsterdam, 1995), vol. 1A of *Handbook of Biological Physics*, pp. 1–61.

[60] A.-S. Smith and U. Seifert, Physical Review E **71**(6), 061902 (2005).

[61] D. A. Fedosov, B. Caswell, and G. E. Karniadakis, Biophys. J. **98**(10), 2215 (2010).

[62] T. Auth and N. S. Gov, Biophys. J. **96**, 818 (2009).

[63] S. Tollis, A. E. Dart, G. Tzircotis, and R. G. Endres, BMC Syst. Biol. **4**(1), 149 (2010).

[64] N. S. Gov and S. A. Safran, Biophys. J. **88**, 1859 (2005).

[65] A. Gray, *Differentialgeometrie* (Spektrum Verlag, Heidelberg, 1994).

[66] G. Gompper and M. Schick, *Self-Assembling Amphiphilic Systems*, vol. 16 of *Phase Transitions and Critical Phenomena* (Academic Press, London, 1994).

[67] S. A. Safran, *Statistical Thermodynamics of Surfaces, Interfaces, and Membranes*, vol. 90 of *Frontiers in Physics* (Westview Press, Boulder, 1994).

[68] D. R. Nelson, T. Piran, and S. Weinberg, *Statistical Mechanics of Membranes and Surfaces*, Jerusalem Winter School for Theoretical Physics (World Scientific, Singapore, 1989).

[69] P. B. Canham, J. Theor. Biol. **26**, 61 (1970).

[70] T. Betz, M. Lenz, J.-F. Joanny, and C. Sykes, Proc. Natl. Acad. Sci. U. S. A. **106**(36), 15320 (2009).

[71] H. G. Döbereiner et al., Phys. Rev. Lett. **91**, 048301 (2003).

[72] A. Zilker, H. Engelhardt, and E. Sackmann, J. Phys. (France) **48**, 2139 (1987).

[73] H. G. Döbereiner, Curr. Opin. Colloid Interface Sci. **5**, 256 (2000).

[74] U. Seifert, Adv. Phys. **46**, 13 (1997).

[75] M. Deserno, Macromol. Rapid Commun. **30**, 752 (2009).

D 2 Surface Patterning

G. A. Vliegenthart
Theoretical Soft Matter and Biophysics
Institute for Advanced Simulation
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Lithography **2**
2.1 Photolithography & electron beam lithography 2
2.2 Soft lithography 3
2.3 Colloidal lithography 7

3 Wrinkling thin elastic films coupled to soft elastic media **10**
3.1 Wrinkling flat elastic films 10
3.2 Spherical geometry 12

4 Conclusions **13**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

In the last 50 years the miniaturisation of mechanical, electronic and optical devices, with applications in consumer products, research and development as well as in medicine has advanced to ever smaller length scales. For the structural downsizing of the relevant features in these technological applications surface patterning provides an important tool. Many new strategies have been developed that allow to functionalise a surface in a well defined, reproducible manner with structural length scales between several tens of nanometers and hundreds of micrometers. Nowadays, patterned surfaces play an important role in the development of technological applications ranging from microelectronics [1], bioelectronics [2, 3], biodiagnostics [4], to microfluidic devices [5]. Among the most important factors for the fabrication and the successful application of these structures are resolution (lateral and height), stability, reproducibility, cost and the complexity of the production process.
Different physical and chemical techniques can be employed to create patterned surfaces. In this chapter I will focus on a couple of techniques that are relevant for Soft Matter science and applications. The methods discussed here can be divided in two categories. The first one considers lithographic or printing techniques whereas the second one utilises the buckling instability of a thin elastic layer to create surface patterns.

2 Lithography

In fine art lithography is a technique that was invented by Alois Senefelder (1771-1834). An image or drawing is transferred from a flat porous stone surface to paper [6]. The underlying mechanism is the immiscibility of oil and water. On a flat piece of stone an image is drawn with a greasy substance, for instance a wax stick. The stone, with image, is then treated with water such that only the surface area that is not covered by the drawing absorbs the water. Subsequently the wet stone is rolled with greasy ink which in turn sticks only to the area where the drawing is put on. The stone with the ink is then pressed on a sheet of paper to make the reproduction. For different colours different stones are required. Since the stone hardly wears during the printing process many copies can be made without or with very little loss of quality. The concept that a selected surface pattern can be reproduced has evolved into modern lithographic techniques such as photolithography, soft lithography and colloidal lithography which will be discussed briefly in the following sections.

2.1 Photolithography & electron beam lithography

Photolithography is the main technology that is used to make integrated circuits. In photolithography [1, 7, 8] a light sensitive photoresist (this is typically a polymer) is spin coated on a substrate, for instance a silicon wafer. The photoresist is subsequently exposed with light using a mask which contains the pattern to be imprinted. The resist is then developed which finalises the process of pattern transfer. The photoresist can be positive or negative meaning that either the area exposed to light can be removed or the area not exposed to light can be removed. The resolution of the patterns to be imprinted is $R = k\lambda/\mathrm{NA}$ with k a constant depending on the photoresist, λ the wavelength of the light used and NA the numerical aperture of the lens. Using very advanced techniques the resolution of photolithography can reach the range of 50-200 nm. For the preparation of a surface a clean room facilities is required. The total number of

processing steps is typically on the order of about ten, containing several cleaning, preparation, developing, baking and post processing steps. This leads to high operational costs and together with the fact that this technique is not suitable for non-planar substrates [9] it is a not very attractive option for biological and chemical applications.
Electron beam lithography uses a focussed electron beam to write directly a pattern on a substrate coated with a photoresist, typically Poly(methyl methacrylate) (PMMA). The mechanism is very similar as in photolithography. The electron beam changes the solubility of the photoresist enabling the selective removal of either the exposed or the non-exposed regions. The main difference is that in electron beam lithography no mask is needed. The resolution of this technique goes down to 15 nm.

2.2 Soft lithography

Soft lithography is concerned with replicating patterns using soft elastomeric stamps, moulds and photomasks [10, 12, 11]. A topographically defined structure (often in a photoresist layer) on a master is replicated in a soft elastomer. This replication process can be carried out under normal laboratory conditions. The main material used is poly(dimethylsiloxane) (PDMS). PDMS is an optically transparent elastomer which can be crosslinked. It's stiffness (Young's modulus) can be tuned from very soft to rigid by varying the amount of crosslinker during preparation. This is important for instance when working with biological cells and tissues where optimal cell growth and cell function rely on a substrate that has elastic properties as close a s possible to the natural environment [13]. Other advantages of PDMS include that it is insulating, it has a low permeability to solvents, it is inert, nontoxic and has a low surface free energy [14, 15]. The latter implies that replicas can easily be released from their moulds. The surface properties of PDMS can be easily modified by oxidising the outer polymer layer.
The soft lithographic technique is (in general) considerably cheaper than photolithography and allows for a better resolution (down to 10 nm). But this depends of course on the quality of the mask. Another important advantage over other techniques is that multiple stamps can be created from one master without loss of quality. The technique can be fast, a typical cycle from idea until the stamp is operational can be less then about 24 hours [10]. Since the first experiments many strategies have been developed for preparing the initial stamp. A schematic example of such a procedure is shown in Fig.1. The height of the topographical features can be adjusted by the photoresist layer thickness as well as by etching.
Soft lithography can also be used to make for instance very well defined arrays of elastomeric micropillars [16, 17] of which an example is shown in Fig.2a. The preparation involves a lithographic mask prepared using electron beam lithography that is used to make a photolithographic image onto a 30μm thick layer of photoresist. The non-crosslinked photo resist was removed and PDMS was dispensed onto the microstructured master and cured [17]. The typical Youngs modulus (stiffness) of the material was in this case about 600 kPa. In this structure cardiac myocytes (heart cells) are grown (Fig.2b) which are able to form a network spanning several pillars. Beating, that is periodic contraction of the cells, then leads to deformation of the pillars and from the known elastic response of the pillars the forces involved can be estimated. These forces turn out to be in the range of 100-200 nN [17].
Nano imprint lithography is a variation of soft lithography where a stamp with a predefined topographical structure is pressed into a soft substrate, typically a softened thermoplastic polymer or a liquid polymer precursor that is subsequently cured by cooling down or irradiation with UV light [8, 18].

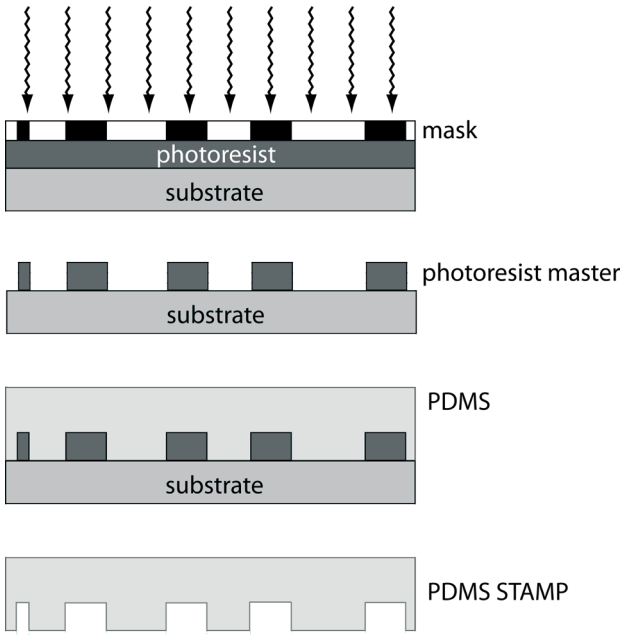


Fig. 1: *Preparation of a stamp. A substrate is coated with photoresist. The photoresist is covered with the mask that is to be reproduced. The photoresist is irradiated and the unexposed photoresist is etched away. PDMS is poured over and cured. The stamp is peeled from the master. Inspired by Ref.[12]*

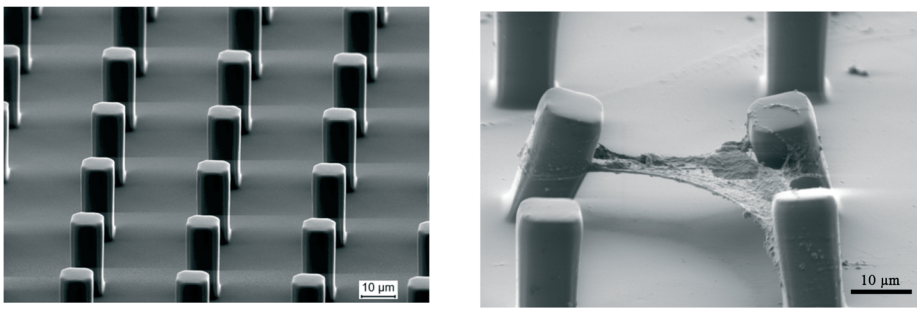


Fig. 2: *Left: Scanning electron micrograph of a micropillar array. Image from Ref.[17] Right: Micropillars deformed by myocytes. Image kindly provided by Nico Hampe (ICS7, Forschungszentrum Jülich).*

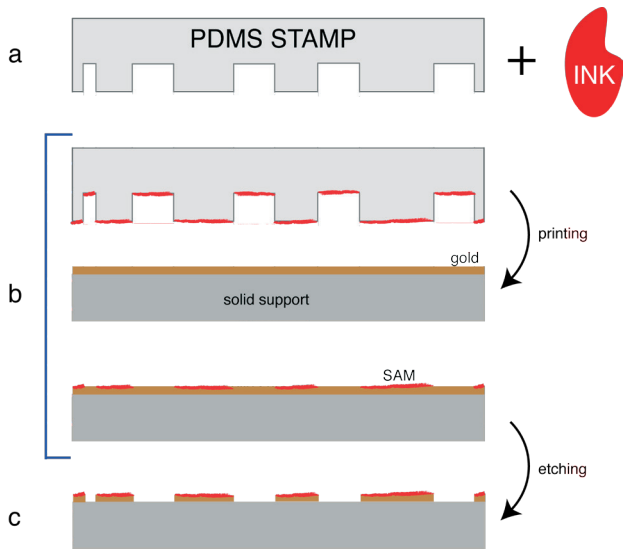


Fig. 3: *The principle of mico contact printing, inspired by Ref.[19]. a) A PDMS stamp is created as in Fig.1; b) the PDMS stamp is dipped into ink (in this case an alkanethiol solution) and stamped onto a gold surface where the thiol groups react covalently with the surface. A self assembling monolayer is formed; c) the unprotected gold surface can be etched away to form the desired topographic structure.*

Micro-contact printing utilises an elastomeric stamp to print a pattern with some kind of ink on a surface. The ink functionalises the surface in a particular way in order to let molecules or larger structures like cells or bacteria selectively adsorb on the surface. The technique of micro-contact printing (μCP) was introduced by Whitesides in the early 1990's [19]. Using a stamp of poly(dimethylsiloxane) (PDMS), they were able to print self assembled monolayers of alkanethiols onto a flat golden surface to form a predetermined pattern. Subsequent treatment of the partially covered surface with a solution of 1M KOH, 0.1 M KCN and dioxygen etches the unprotected parts of the surface [19, 20]. This procedure is illustrated in Fig.3. Since the invention of μCP the number of inks has increased enormously [21] and in principle any substance that has a higher affinity for the surface than for the stamp can be printed. Most conveniently the ink undergoes some chemical reaction with active surface groups available [22, 23] which leads to a monomolecular layer. If the surface bound molecules contain some functional or reactive group other molecules, particles functionalised with complementary interactions or even whole cells or bacteria can be adsorbed specifically.

Using μCP, patterns with typical dimensions of one to several hundreds of micrometers can be obtained. There are however some technical problems that arise from the deformability of the stamp, a problem that mainly appears for high aspect ratio structures (width/height of the pilars) [3, 24]. This is shown in Fig.4. Moreover, swelling of the stamp due to absorption of solvent also changes the shape and reduces the precision of the pattern. The problem of deforming the stamp when pressing it on the surface can be partially overcome by so-called submerged μCP (SμCP) where the printing is performed under water (submerged) and the water stabilises the stamp [24].

An important application of μCP in cell biology [25, 26], tissue engineering [27] and also the

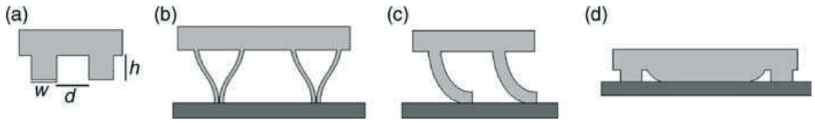


Fig. 4: *The aspect ratio and the spacing of the structures on an elastomeric stamp strongly affects their mechanical stability and the accuracy of the print. Reproduced from Ref.[**?**].*

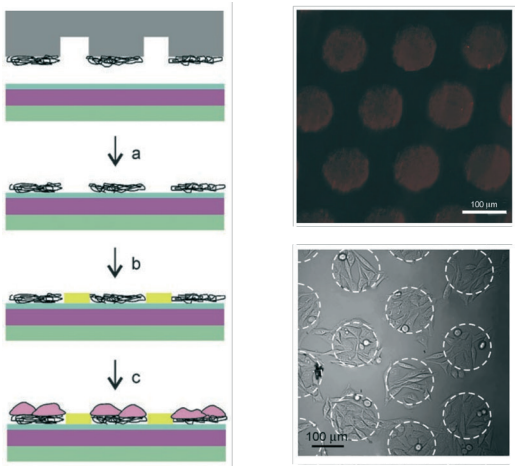


Fig. 5: *Left: a) The collagen-like protein Col3a1 (coil-like structure on the stamp) is printed on a functionalised surface using an oxidised PDMS stamp. b) The remaining area of the surface is covered with Methoxypoly(ethylene glycol)amine to form areas that are resistant to cell adhesion. c) HeLa cells bind selectively on the Col3a1 proteins. Right: d) Confocal microscopy image of printed Col3a1 patterns (100-µm dots separated by 100 µm) labeled with the fluorescent dye Lissamine rhodamine B. e) Patterns of HeLa cells obtained by µCP of protein Col3a1. Figure reproduced from Ref.[31]*

development of biosensors [28] is that of patterning proteins [26, 29, 30]. The surface is first functionalised such that proteins can bind to it and subsequently an ink containing proteins is stamped on it. This structured pattern of proteins can be used for instance to study the adhesion and movement of cells on a substrate.

In Fig.5 we show preferential cell adsorption on a surface that is patterned with a specific protein [23, 31]. A surface is coated with an aldehyde terminated self assembled monolayer. On this layer a pattern of collagen-like proteins is printed using an oxidised PDMS stamp. Oxidation renders the PDMS surface hydrophilic. The uncovered regions were made nonadhesive by adsorbing Poly(ethylene glycol) molecules. Finally, human malignant carcinoma (HeLa) cells were seeded and incubated on the surface [31]. For successful localisation of cells the size and spacing of adsorption sites is important. If the distance between the sites is too small, cells will also occupy the space between the adsorption sites. For this particular system is turns out that the pattern evolves in time. After incubation for more than about 72 hours cells start to spread also to regions outside the adsorption sites. This is possibly due to deliberate modification of the interstitial regions by by the cells by secreting degrading enzymes, applying local pH changes or applying mechanical stresses [31, 32].

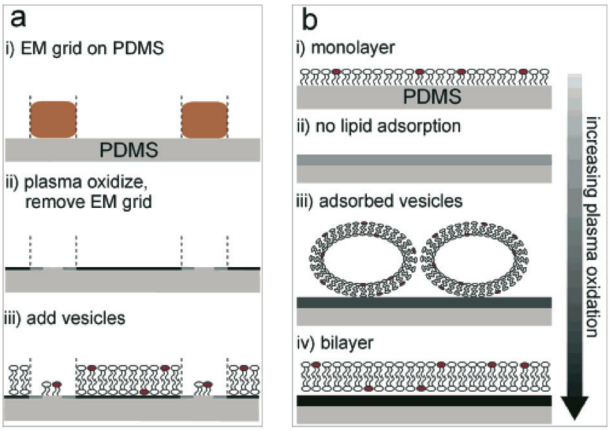


Fig. 6: *a-i) An electron microscopy grid is used to cover a PDMS substrate. a-ii) The unprotected area is plasma oxidised. The increasing grayscale indicates increasing plasma oxidation. a-iii) Vesicles are added that adsorb to form monolayers on the bare hydrophobic PDMS surface and bilayers on the hydrophilic areas. b) Schematic illustration of the dependence of the supported lipid phase on the plasma oxidation time. (i) Fluid supported monolayers form on unmodified PDMS,while (ii) a small degree of plasma modification results in a surface that rejects lipid assembly (gap region). (iii) Further plasma oxidation of the surface causes vesicles to adsorb, ultimately leading to (iv) formation of a fluid supported bilayer. Figure reproduced from Ref.[33]*

Lenz and coworkers employed another strategy to create a periodic hydrophobic-hydrophylic surface pattern on which bilayers can be selectively adsorbed [33]. A PDMS layer was covered using an electron microscope grid (Fig.6a-i) as a mask and subsequently exposed to an air plasma leading to a local oxidation of the PDMS layer in the uncovered regions (Fig.6a-ii). By varying the plasma cleaning time the extend of oxidation can be controlled as is indicated by the light gray (hydrophobic) and dark gray (hydrophobic) regions [33]. By varying the plasma oxidation time the surface can be modified in such a way that it adsorbs lipid monolayers (no oxidation, hydrophobic surface, Fig.6b-i), no adsorption whatsoever (mildy hydrophilic surface, Fig.6b-ii) it adsorbs vesicles (moderately oxidised, hydrophilic surface but not attractive enough to destroy vesicles, Fig.6b-iii) or bilayers adsorb (strongly oxidised, hydrophilic surface, Fig.6b-iv).

2.3 Colloidal lithography

Colloidal lithography provides another way of creating surface patterns and is based on the principle of colloidal crystallisation in two dimensions [34, 35]. For a mono disperse system this leads to an ordered monolayer of colloidal particles. These colloidal particles are in the size range of 50 nm to several micrometers and large enough to undergo Brownian motion. The monolayer which can also consist of more than one size of colloidal particles can be either used directly as it is or it can be used as a mask similar to what we have seen in the previous section. A variety of modifications like metal deposition and etching steps can be then applied to obtain patterned surfaces. The main advantage of colloidal lithography over soft lithography

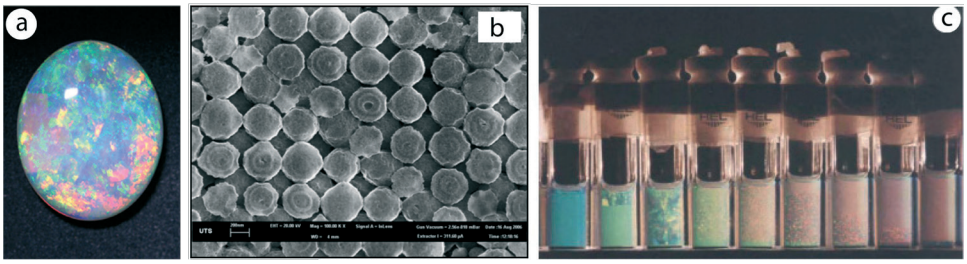


Fig. 7: *a) natural opal, image reproduced from http://www.southshorediamond.com; b) cubic lattice of silica spheres of a Coober Pedy opal, image reproduced from Ref.[38]; c) suspension of synthetic colloidal hard spheres undergoing crystallisation, image reproduced from Ref.[40].*

is that relatively simple synthetic methods can be used to make the particles and that one can completely circumvent photolithography or electron beam writing for creating masks.

Colloidal crystallisation Colloidal crystals appear in nature as opals [36, 37, 38] (Fig.7a,b) which are well known for their beautiful iridescent colors. If they are put under a microscope a three-dimensional ordered structure of small spherical particles is observed. These particles were formed first by dissolution of silica by chemical weathering from soluble silicates and subsequent nucleation of silica particles in this solution at an appropriate pH [38]. Sedimentation and compression then lead to the formation of crystals. The silica particle size in opals is typically between 150-300 nm. Ordering of particles of this size in a regular lattice gives rise to diffraction and Bragg reflections in the visible part of the electromagnetic spectrum so that when we use our head as a crystallographic goniometer different colors are observed. The formation of synthetic opals, colloidal crystals, has been studied extensively over the last 30 years. It is now known that if hard sphere-like colloidal particles are prepared with a size polydispersity lower than about 10% they can, under the appropriate conditions, that is at large enough volume fractions, undergo a phase separation into a fluid and a solid phase, a colloidal crystal. The underlying crystallisation transition was already predicted in the early 1950's by computer simulations of hard spheres [39]. Many years later this phase transition was for the first time extensively experimentally investigated using synthetic colloidal particles [40]. It turns out that in bulk suspensions the initially isotropic fluid phase becomes unstable with respect to the crystalline phase at volume fractions of about 50% and phase separates into a colloidal fluid coexisting with a colloidal crystal whereas above 55% the system is completely crystalline. The transition from colloidal fluid to colloidal crystal is nicely seen in the sequence of images in Fig.7c. Note that in the absence of attractive interactions a colloidal crystal is formed by densification of the colloidal fluid, thats is by purely changing the entropy. At low densities the colloidal fluid is stable and fully characterised by *translational entropy* of the particles. When increasing the density the translational entropy decreases as colloidal particles increasingly hinder each other. At the transition which for hard spheres happens around 50% in volume of colloidal particles it becomes more favourable for the particles to order in a lattice and gain vibrational degrees of freedom around average lattice positions. The colloidal crystallisation transition is not limited to only one-component systems or spherical particles. Moreover, for anisotropic particles positional and orientational order play a role and for mixtures of for instance colloidal spheres and colloidal rods or colloidal plates and colloidal spheres the phase behaviour can become rather complicated.

Colloidal crystals can also be obtained in two dimensions even though in two dimensions long ranged positional order is absent at finite temperatures[41, 42].

Monolayer assembly There are many experimental protocols to obtain monolayer or two-dimensional colloidal crystals in an experimental setting. A comprehensive overview of methods and their advantages and disadvantages is given by Vogel [34]. Ye and Qi in addition also discuss several applications of monolayer crystals [43]. Here I will briefly discuss a few preparation methods and some applications.
The simplest method to produce a colloidal crystal is to use a gravitational field to precipitate particles on a wall [44]. This method is slow and one has very little control over the number of layers that is grown [35].
A well documented and frequently used preparation method utilises a combination of evaporation induced forces and capillary forces between colloidal particles. A drop of a colloidal suspension is placed in a teflon ring on a substrate and the solvent is slowly evaporated. The teflon ring prevents the particles to accumulate at the boundary (as in a drying coffee drop) and induces crystal nucleation at the center when the concentration is large enough. The evaporation leads to a particle flux from the boundary of the droplet radially inwards which causes crystallite to grow [45, 46, 47].
Vogel recently developed a fast and simple method, called direct assembly, not requiring any special equipment, to create high quality monolayers of spherical particles [34, 48]. The method resembles the one of Retsch and co-workers in which particles first are put in a disordered layer on a flat surface which is subsequently immersed in a host fluid under a slight angle. The particles form then a dense monolayer at the interface. This monolayer is picked up by immersion of the desired substrate under a slight angle and lifting it from the fluid [49].
Vogel's method is illustrated in Fig.8 and consists of the following steps. First, colloidal particles are added to the interface until the interface is fully covered. Adding too many particles leads to buckling of the monolayer and to multilayer formation. Then, a substrate, which can be flat or have any topographical structure, is immersed in the liquid and lifted from the subphase at a shallow angle. The substrate is then dried and the monolayer ready to be used [48]. It turns out that the solvent quality and the addition of small amounts of surfactant have a large effect on the structure of the monolayer. In particular, monolayers obtained in pure water (milliQ grade) turn out to be very inhomogeneous. The addition of small amounts of surfactant leads to a soft barrier on the interface that enhances the packing of the crystal. Moreover, negative surfactant molecules adsorbing on the particles increase the charge of the particles leading to an improved stabilisation of the particles against attractive van der Waals and capillary interactions in turn leading to improved crystallinity [34, 48, 50]. Structures different from the standard hexagonal packing of particles can be obtained by deposition of particles on pre-structured surfaces like in lines [51] or a square lattice of patterned nano wells [52].
Post processing of colloidal monolayers can be used to create a large variety of possible structures including nano-nets, nano-bowls, nano-pillars, nano-cones, etc. [43, 50] and references therein.

Binary colloidal crystals Like in atomic systems where crystals can be composed of different sized atoms, colloidal crystals can be grown with for instance particles of different size, different shape or different composition (chemistry). In fact, in nature it turns out that certain Brazilian opals are composed of two sizes of colloidal particles [53, 54]. The mixing of two

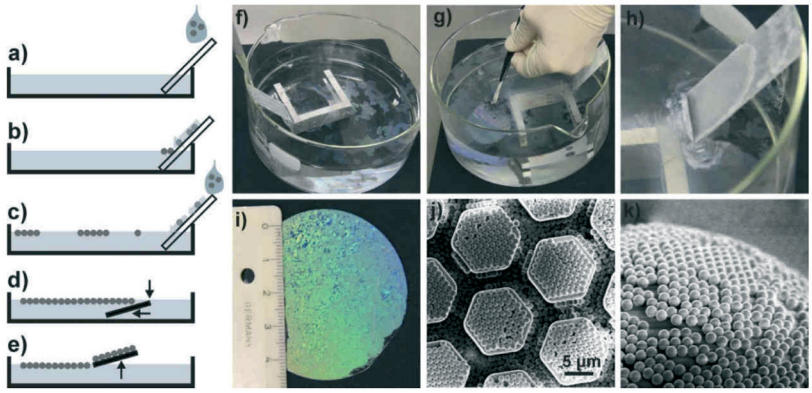


Fig. 8: *a,b) Particles are added to the interface using a glass slide; c) a close-packed monolayer is formed; d,e) the monolayer is transferred to the desired substrate; f,g) transfer to wafer; h) dried monolayer; i) macroscopic crystalline domain; j) particles assembled on micropillar array; k) particles assembled on large sphere. Reproduced from Ref.[55]*

sizes of colloidal particles gives yet another relatively easy route to create more complicated surface patterns. The important parameters to tune the structures are the size ratio between large and small particles and the stoichiometry, the relative concentrations of the particles. Figure 9 shows various possible structures for large colloidal particles crystallised in a hexagonal lattice with small particles located at the interstitial sites. The exact structures depend on the particle size ratio and the stoichiometry [55]. The main observation is that the interstitial occupancy increases with the particle number ratio of small to large particles R_N. Similar crystal structures as shown in Fig.9 were also reported in Refs.[54, 56].

3 Wrinkling thin elastic films coupled to soft elastic media

When an apple is left to dry it will show after some time a wrinkled surface pattern (see Fig.10a). This pattern results from a simple effect: the apple skin which upon drying retains its surface area has to buckle in order to accommodate the decreasing volume of the drying apple [57, 58]. This effect, where a thin surface layer coupled to an elastic substrate buckles due to the compression of the elastic substrate, leads to wrinkling patterns that can have long range order on flat substrates [59, 60, 61, 62, 63]. An example of such a wrinkling pattern, in this case a disordered one, is shown in Fig.10b.

3.1 Wrinkling flat elastic films

The wrinkling of flat elastic films has been studied extensively over the last 20 years [57, 59, 60, 61, 62, 64, 66, 67, 68]. Systems undergoing a wrinkling transition can be prepared experimentally in a sequence of steps. For instance, a block of PDMS is heated such that it expands. In the expanded state it is covered by a thin metal layer. Subsequent cooling of the system to normal temperatures leads to compression of the bulk PDMS to its original volume which in turn leads to a lateral compression of the top layer that prefers the expanded size. For large enough stresses the top layer forms a pattern of wrinkles [59]. The procedure has been repeated

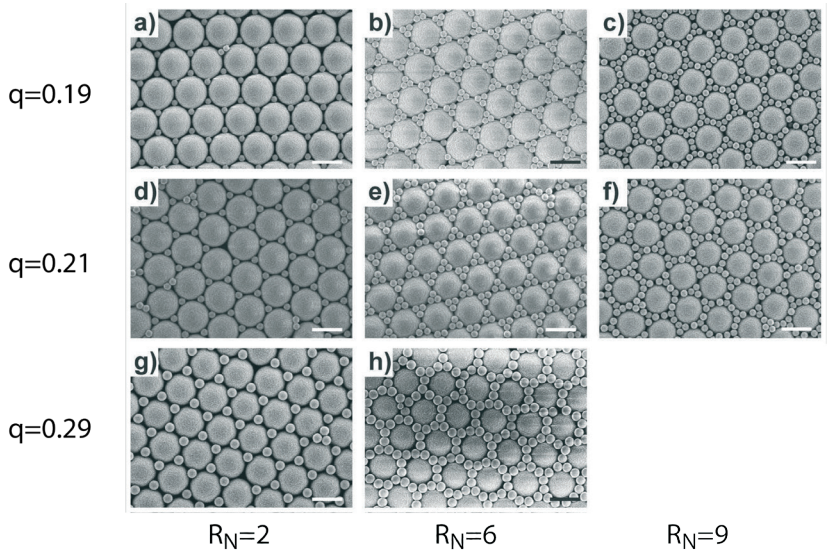


Fig. 9: *Structural changes when adjusting the size ratio $q = d_s/d_l$ and particle number ratio $R_N = N_s/N_l$. The large particles have a diameter d_l =1063 nm. The scale bar is 1μm unless stated differently. Reproduced from Ref.[55]*

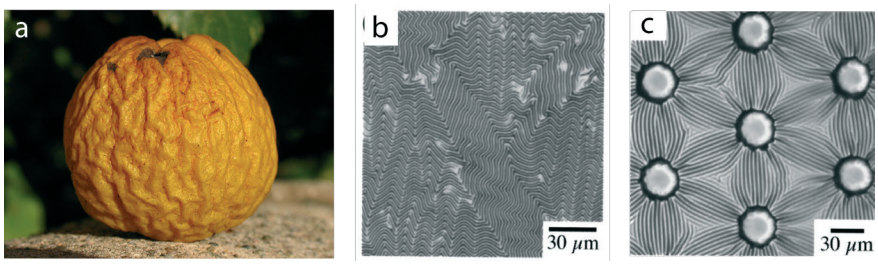


Fig. 10: *a) A drying apple with strongly wrinkled surface. b) The wrinkling pattern of PDMS far away from edges or surface topographical structures. Here, many partially ordered domains are found. c) Ordered wrinkle pattern on a structured surface. Reproduced from Ref.[60]*

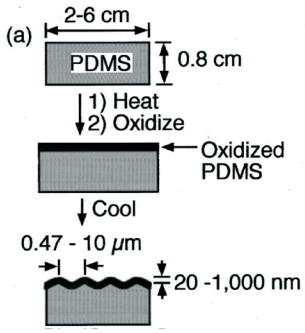


Fig. 11: *A solid block of PDMS with known stiffness E_s and Poisson ratio ν_s is heated to expand it and subjected to an air plasma to oxidise the top layer. The sample is then cooled and wrinkles with a characteristic wavelength appear. Reproduced from Ref.[60]*

for various metals confirming the mechanism of wrinkle formation. Moreover, upon re-heating and expansion the wrinkles disappear whereas after cooling the wrinkles re-appear. The process is thus fully reversible.

Another procedure is to heat the PDMS block and to expose the top layer to an air plasma. This oxidises the top layer by which its elastic properties change. Cooling down again leads to the formation of wrinkles [60]. By changing the plasma exposure time the thickness and the stiffness of the top layer can be modified. Moreover, the expansion (and thus compression when cooling) can be tuned by changing the temperature [60]. The patterns that are found are for these systems disordered (Fig.10b) as long as the surface is homogeneous whereas patterning by bas relief (inducing step-like features) leads to ordering of the wrinkles at the edges of the steps (Fig.10c).

The scaling of the typical wavelength λ of the wrinkles can be obtained from a balance of the bending energy of the stiff surface layer and the stretching energy of the substrate and is given by [59, 62, 66],

$$\lambda \sim C\, h \left(\frac{E_p}{E_s}\right)^{1/3}. \tag{1}$$

With h the thickness of the top layer, E_p the stiffness (Youngs modulus) of the top layer, E_s the stiffness of the substrate layer and C a factor that still contains the Poisson ratios of the top layer and the substrate as well as a numerical pre-factor.

This shows that in fact bending the stiff surface prefers large wavelength wrinkles whereas compression of the substrate prefers short wavelength wrinkles. The competition of the two leads to intermediate wavelength wrinkles [57].

3.2 Spherical geometry

As illustrated by the drying apple, pattern formation by wrinkling is certainly not limited to flat geometries. In fact, pattern formation on a spherical topology has some interesting consequences. Most importantly, long range crystalline order, a perfect organisation of features on the spherical surface is not possible. The packing of equal discs on a spherical surface requires an excess of twelve 5-fold disclinations. A nice example of such a packing is found in the protein shell structure of spherical viruses [69]. Li and coworkers studied the deformation of

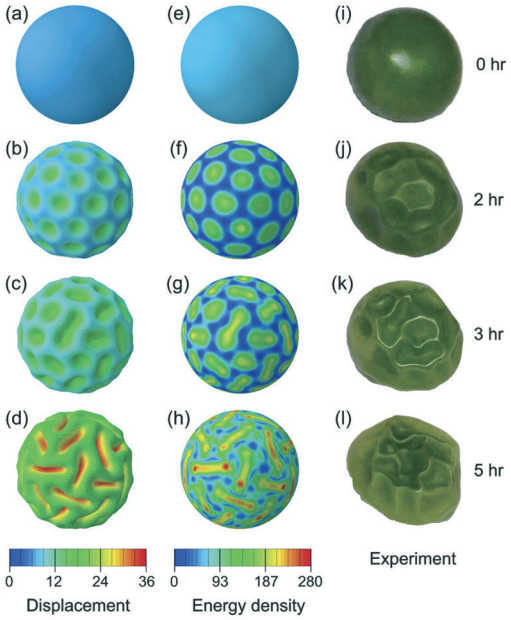


Fig. 12: *Buckling of a core shell particle subject to isotropic compression as a function of time. The ratio of the shear moduli of the core and shell is* $\mu_c/\mu_s = 10$*, the core radius* $R_c = 100$ *and the shell thickness* $h = 4$*. Reproduced from Ref.[58]*

core-shell elastic spheres numerically using finite element calculations. Figure 12 shows the structures from calculations compared to drying experiments on green peas. Initially the core shell particle is isotropically compressed and retaining its spherical shape. After initial stress condensation a pattern of closed packed small amplitude perturbations is found similar as in [70]. This resembles the pattern found at the Föppl-von-Kármán instability for pressure induced buckling of a spherical shell. For the latter case, one of the initially formed indentations grows to form the single indented ground state of a buckled shell [70, 71]. In the case discussed here, the solid elastic core prevents the asymmetric bowl-shaped ground state of a deflated shell to be formed and instead stabilises a homogenous distribution of small indentations or folds depending on the compression of the particle[58].

4 Conclusions

We have discussed three different methods to make patterned surfaces. Soft lithographic methods are fast, reproducible and versatile. However, often one still relies on masks created by photolithography or electron beam writing. Colloidal lithography is based on two-dimensional crystallisation of colloidal spheres and provides a cheap and relatively easy route to create many different nanostructures with underlying lateral crystalline ordering. Finally, the wrinkling of thin elastic films can create ordered and disordered periodic patterns on flat substrates as well as weakly deformed regular and highly deformed and crumpled structures on curved surfaces.

Acknowledgement

I would like to thank Prof. Rudolph Merkel and Nico Hampe from the Institute of Complex Systems ICS-7, Biomechanics, for sharing their expertise on Soft Lithography with me.

References

[1] R.C. Jaeger, "Lithography" Introduction to Microelectronic Fabrication (2nd ed.). Upper Saddle River: Prentice Hall. (2002).

[2] G.W. Gross, B.K. Rhoades, H.M.E. Azzazy, M.C. Wu, Biosens. Bioelec. **10** 553 (1995).

[3] L. Lauer, C. Klein, A. Offenhäusser, Biomaterials **22** 1925 (2001).

[4] G. MacBeath, S.L. Schreiber, Science **289** 1760 (2000).

[5] C. Rivet, H. Lee, A. Hirsch, S. Hamilton, H. Lu, Chem. Eng. Sci. **66** 1490 (2011).

[6] A. Senefelder, *The invention of lithography* New York: The Fuchs and Lang Manufacturing Company (1911).

[7] S. Okazaki, J. Vac. Sci. Technol. B **9** 2829 (1991).

[8] Z. Nie, E. Kumacheva, Nat. Mat. **7** 277 (2008).

[9] W.D. Deninger, C.E. Garmer, J. Vac. Sci. Technol. B **6** 337 (1988).

[10] Y. Xia and G. M. Whitesides, Angew. Chem., Int. Ed. **37** 550 (1998).

[11] B.D. Gates, Q. Xu, M. Stewart, D. Ryan,C. Grant Willson, G.M. Whitesides, Chem. Rev. **105** 1171 (2005).

[12] T. Deng, H. Wu, S.T. Brittain, G.W. Whitesides, Annal. Chem. **72** 3176 (2000).

[13] J.N. Lee, X. Jiang, D. Ryan, G.M. Whitesides, Langmuir **20** 11684 (2004).

[14] J.C. McDonald, G.M. Whitesides, Acc. Chem. Res. **35** 491 (2002).

[15] Y. Fainman, L. Lee, D. Psaltis, C. Yang, *Optofluidics: fundamentals, devices and applications* McGraw-Hill (2009).

[16] S. Ghassemi, N. Biais, M. Maniura, S.J. Wind, M.P. Shettz, J. Hone, J. Vac. Sci. Technol. B **26** 2549 (2008).

[17] A. Kajzar, C. M. Cesa, N. Kirchgener, B. Hoffmann, R. Merkel, Biophys. J. **94** 1854 (2008).

[18] L. J. Guo, Adv. Mater. **19** 495 (2007).

[19] A. Kumar, G.W. Whitesides, Appl. Phys. Lett. **63** 2002 (1993).

[20] A. Kumar, H.A. Biebuyk, N.L. Abbott, G.W. Whitesides, J. Am. Chem. Soc. **114** 9188 (1992).

[21] S.A. Ruiz, C.S. Chen, Soft Matter **3** 1 (2007).

[22] J.M. Spruell, *The Power of Click Chemistry for Molecular Machines and Surface Patterning* Springer Theses (2011).

[23] C. Wendeln, B.J. Ravoo, Langmuir **28** 5527 (2012).

[24] F. Bessueille, M. Pla-Roca, C.A. Mills, E. Martinez, J. Samitier, A. Errachid, Langmuir **21** 12060 (2005).

[25] Y. Ito, Biomaterials **20** 2333 (1999).

[26] R.S. Kane, S. Takayama, E. Ostuni, D.E. Ingber, G.M. Whitesides, Biomaterials **20** 2363 (1999).

[27] N. Gadegaard, E. Martines, M.O. Riehle, K. Seunarine, C.D.W. Wilkinson, Microelectron. Eng. **83** 1577 (2006).

[28] D. Schwaab, Ph.D. Thesis, RWTH Aachen, Fakultät für Mathematik, Informatik und Naturwissenschaften, Germany, (2007).

[29] A.S. Blawas, W.M. Reichert, Biomaterials **19** 595 (1998).

[30] S.R. Coyer, A.J. Garcia, E. Delamarche, Angew. Chem. Int. Ed. **46** 6837 (2007).

[31] D.I. Rozkiewicz, Y. Kraan, M.W.T. Weten, F. A. de Wolf, V. Subramaniam, B.J. Ravoo, D.N. Reinhoudt, Chem. Eur. J. **12** 6290 (2006).

[32] C.M. Nelson, S. Raghavan, J.L. Tan, C.S. Chen, Langmuir **19** 1493 (2003).

[33] P. Lenz, C.M. Ajo-Franklin, S.G. Boxer, Langmuir **20** 11092 (2004).

[34] N. Vogel, *Surface Patterning with Colloidal Monolayers* Springer Theses (2011).

[35] Ye Yu and Gang Zhang, Colloidal Lithography, Updates in Advanced Lithography, S. Hosaka (Ed.) (2013).

[36] J.B. Jones, J.V. Sanders, E.R. Segnit, Nature **204** 990 (1964).

[37] J.V. Sanders, Am. Min. **60** 749 (1975).

[38] A.G. Smallwood, P.S. Thomas, A.S. Ray, J. Austr. Cer. Soc. **44** 17 (2008).

[39] B.J. Alder, T.E. Wainwright, J. Chem. Phys. **27** 1208 (1957).

[40] P. Pusey, W. van Megen, Nature **320** 340 (1986).

[41] R. Peierls, Ann. Inst. Henri Poincaré **5** (1935).

[42] N. Mermin, H. Wagner, Phys. Rev. Lett. **17** 1133 (1966).

[43] X. Ye, L. Qi, Nano Today **6**, 608 (2011).

[44] S.H. Park, D. Qin, Y. Xia, Adv. Mater. **10** 1028 (1998).

[45] N.D. Denkov, O.D. Velev, P.A. Kralchevsky, I.B. Ivanov, H. Yoshimura, K. Nagayama, Langmuir **8** 3183 (1992).

[46] P.A. Kralchevsky, K. Nagayama, Langmuir **10** 23 (1994).

[47] P.A. Kralchevsky, K. Nagayama, Adv. Coll. Interface Sci. **85** 145 (2000).

[48] N. Vogel, S. Goerres, K. Landfester, C.K. Weiss, Macromol. Chem. Phys. **212** 1719 (2011).

[49] M. Retsch, Z. Zhou, S. Rivera, M. Kappl, X.S. Zhao, U. Jonas, Q. Li, Macromol. Chem. Phys. **210** 230 (2009).

[50] C. Li, G. Hong, P. Wang, D. Yu, L. Qi, Chem. Mater. **21** 891 (2009).

[51] X. Yan, J.M. Yao, G. Lu, X. Li, J.H. Zhang, K. Han, B. Yang, J. Am. Chem. Soc. **127** 7688 (2005).

[52] N.N. Khanh, K.B. Yoon, J. Am. Chem. Soc. **131** 14228 (2009).

[53] S. Hachisu, S. Yoshimura, Nature **283** 188 (1980).

[54] K.P. Velikov, C.G. Christova, R.P.A. Dullens, A. van Blaaderen, Science **296** 106 (2002).

[55] N. Vogel, L. de Viguerie , U. Jonas , C. K. Weiss, K. Landfester, Adv. Funct. Mater. **21** 3064 (2011).

[56] V. Kitaev, G.A. Ozin, Adv. Mater. **15** 75 (2003).

[57] E. Cerda, L. Mahadevan, Phys. Rev. Lett. 90 074302 (2003).

[58] B. Li, Y-P Cao, X-Q Feng, H. Gao, Phys. Rev. Lett. **106** 234301 (2011).

[59] N. Bowden, S. Brittain, A. G. Evans, J. W. Hutchinson, G. M. Whitesides, Nature **393** 146 (1998).

[60] N. Bowden, W.T.S. Huck, K.E. Paul, G.M. Whitesides, Appl. Phys. Lett. **75** 2557 (1999).

[61] J. Groenewold, Physica A **298** 32 (2001).

[62] J. Genzer, J. Groenewold, Soft Matter **2** 310 (2006).

[63] E. Cerda, K. Ravi-Chandar, L. Mahadevan, Nature **419** 579 (2002).

[64] X. Chen, J.W. Hutchinson, J. Appl. Mech. **71** (2004).

[65] B. Li, Y-P Cao, X-Q Feng, H. Gao, Soft Matter **8** 5728 (2012).

[66] H.G. Allen, Analysis and design of structural sandwich panels (Pergamon, New York, 1969).

[67] M.A. Biot, Proc. Roy. Soc. London, Ser A, **242** 444 (1957).

[68] M-W. Moon, S.H. Lee, J-S Sun, K.H. Oh, A. Vaziri, J.W. Hutchinson, PNAS **104** 1130 (2007).

[69] J. Lidmar, L. Mirny, D.R. Nelson, Phys. Rev. E **68** 051910 (2003).

[70] J. Paulose, D.R. Nelson, Soft Matter **9** 8227 (2013).

[71] G.A. Vliegenthart, G. Gompper, New J. Phys. **13** 045020 (2011).

D 3 Nanocomposites

G. J. Schneider
Jülich Centre for Neutron Science & Institute of Complex Systems
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Preparation of Nanocomposites **6**
2.1 Melt Intercalation 7
2.2 Intercalation of Polymer from Solution 7
2.3 *In Situ* Intercalative Polymerization Method 7

3 Properties of Nanocomposites **7**
3.1 Mechanical Properties 7
3.2 Heat Distortion Temperature 9
3.3 Thermal Stability and Fire Retardant Properties 10
3.4 Gas Barrier Properties 11
3.5 Optical Transparency 13
3.6 Conclusions 13

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Modern life depends strongly on polymer based composites. From simple shoes to hi-tech vessels and electronics, many of them are based on polymers in a mixture with other substances, such as inorganic particles. Actually, composites are not a new invention. They can be traced back to ancient times when evolution led to first composites. In some cases, organic materials, like proteins, and inorganic components, like clays, assemble to mixtures that are ductile and possess high tensile-strengths [1]. Famous representatives are nacre, mollusk shells, bones and teeth tissues.

Our bones are very good examples of composite materials. They are composed of the protein collagen in the form of elongated fibrils, with three of them are twisted to a helix. Within the protein, tiny apatite crystallites, size $\sim 2-4$ nm, are aligned parallely to these collagen helices. Up to now, there is no detailed understanding, how these natural materials are produced [2, 3]. Due to their favorable properties that cannot be achieved using neat materials, assembling artificial composites is quite obvious.

The first artificial composites seem to trace back to the ancient Maya. They painted their walls, potteries and created art (Fig. 1) by using the so-called Maya blue, a mixture of indigo and palygorskite, sometimes called attapulgite, a magnesium aluminum phyllosilicate [1]. Most impressingly, the color kept its brilliance even after several centuries despite of heat, humidity and radiation from the sun.

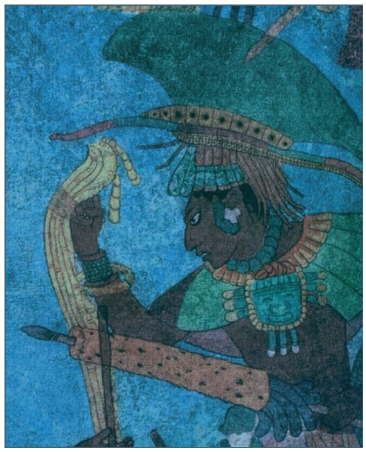


Fig. 1: *Mayan warrior from Bonampak. (Reproduced from Ref. [4])*

At first glance, keeping such art might be nice but unimportant for survival of mankind. However, up to now, scientists are interested in Maya blue, due to its resistance against chemicals, solvents, oxidizing and reducing substances [5, 6]. Obviously, the protection capability rests on intercalating pigment molecules in the channels of the silicates that act like a cage. Furthermore, indigo seems to react with the clay and changes the color from blue to turquoise.

Nowadays, a variety of composite materials exists, e. g. in skin care and - protection products, in lasers, sensors, liquid crystal displays, scratch resistant coatings, car tires, soccer balls. For example, soccer balls require high levels of impermeability to gases and liquids, because air retention is very important, at least for those 90 min of standard games. Though the number of composites is continuously growing, in many cases, even up to now, their development is based on alchemy or on recipes found by try and error [1].

Different scientists believe the high tensile strength of composites is a consequence of adding nanoparticles with small diameters (< 100 nm). If at least one component has a size within the nanometer range the material is called nanocomposite [8]. Most importantly, the properties of the composite are not only the sum of the individual properties of the components. Obviously, due to mutual interactions, novel features emerge resulting in unprecedented properties.

Many inorganic materials have dimensions at the nanoscale, but more or less belong to a different class called microcomposites. For example, a tire tread is a silica rubber composite. The silica particles usually used have diameters less than 20 nm, well covered by the above definition, thus at the first glance a showcase example for a nanocomposite. However, experiments demonstrate that silica nanoparticles can agglomerate eventually and even connect by chemical bonds [9]. Therefore, instead of the particle diameter, the cluster size might determine the properties. Nevertheless, strong reinforcement, i. e. a strong increase of the modulus, measured by dynamical mechanical experiments is observed in such materials [10]. Thus, distinguishing different types of composites simply by one definition might be questionable, related to the plethora of phenomena that could occur. If not explicitly mentioned, hereafter, emphasis is put on the class of true nanocomposites, i. e. we assume one component, e. g. the inorganic nanoparticles, has a dimension of less than 100 nm.

Nanoparticles consist of a few to several hundred atoms/molecules and their diameters lie between those of atoms/molecules and the bulk-material. Such an intermediate size leads to interesting effects. Due to their small diameters, the specific surface area is very large, i. e. many atoms are at the surface of the nanoparticles. These atoms are in a higher energetic state, since they are not fully coordinated in comparison with inner particles. Therefore, they exhibit a higher chemical reactivity and catalytic activity [11]. Furthermore, due to the low number of atoms the number of electrons is low. The spacing between single electron levels becomes larger and one falls back to discrete energy levels, instead of quasi continuous bands in solid state materials. Therefore, new properties emerge that are different to those in classical condensed matter, e. g. concerning optoelectronic and magnetic properties of nanoparticles [11, 12].

Hereafter, we focus on the material improvement exploiting the high specific surface area of nanoparticles that are related to a high number of chemical bonds. For example, composites based on the polymer Nylon-6 and the silicate Montmorillonite (MMT) have been extensively studied [13]. In order to test, whether these particles belong to the class of nanocomposites, structural aspects need to be taken into account. Fig. 2 presents some basic ideas on possible morphologies or assemblies of nanoclays in the polymer. On the upper right hand side, the chemical structure is presented. Generally, two types are found, tetrahedral-substituted and octahedral substituted. In the first case, negative charges can be found on the surfaces. By that means, the polymer chains can interact with the surface.

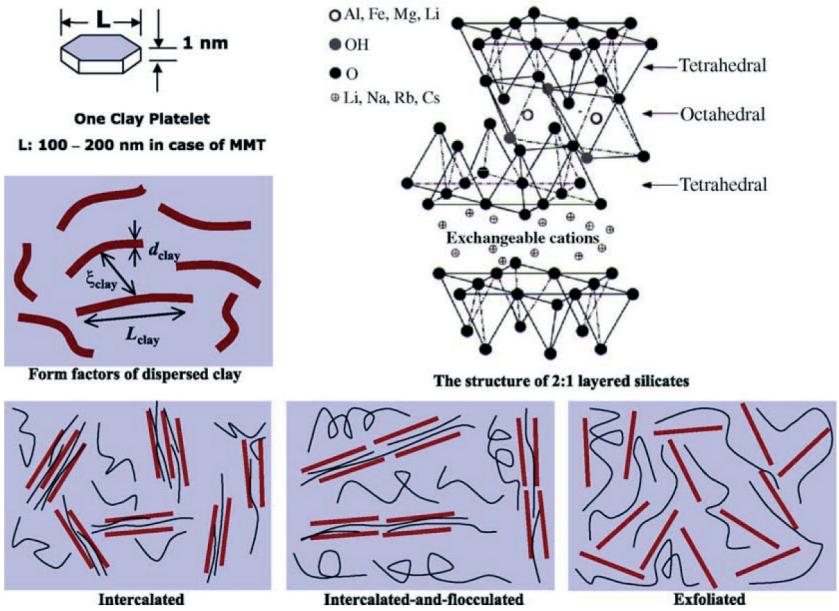


Fig. 2: *Scheme of different structures of the silicate in polymer-silicate nanocomposites. (Reprinted with permission from [13]. Copyright 1997 Elsevier)*

The scheme already suggests that silicates form layers (platelets). One of the most important fact about these sheets is the high ratio between the diameter $L(= L_{\text{clay}})$ and the thickness. Fig. 2 indicates diameters $L_{\text{clay}} \sim 100 - 200\,\text{nm}$, and a thickness $D \sim 1\,\text{nm}$. As a third parameter the distance ξ_{clay} between the sheets plays an important role, e. g. representing the capability to intercalate polymer chains between two layers. The intercalated scenario is displayed on the left hand side (bottom). There, the silicate form stack like assemblies that are well separated from each other or stick together (floc), a process that is called flocculation (middle). Furthermore, due to the capability of the layers to disassemble a so-called exfoliated structure, sometimes called delaminated structure, may form (right).

In order to evaluate these different scenarios experimental results are important, e. g. transmission electron microscopic (TEM) images displayed by Fig. 3. These verify the appearance of all of the different structures postulated by Fig. 2. To reveal details on the local morphology, TEM experiments are very well suited. For example, from such images we easily obtain the typical distances in the samples. (Already, from the overview images presented by Fig. 3, we derive typical distances less than 10 nm between our silicate particles.)

In order to obtain a statistic picture wide-angle X-ray scattering (WAXS) complements TEM very well and is commonly used for that purpose [13]. Basics can be found in Chapter C1 – Scattering of the present textbook. Fig. 4 presents an overview about the typical signatures of the different structures. Furthermore, it indicates that the typical distances are related with the peak positions.

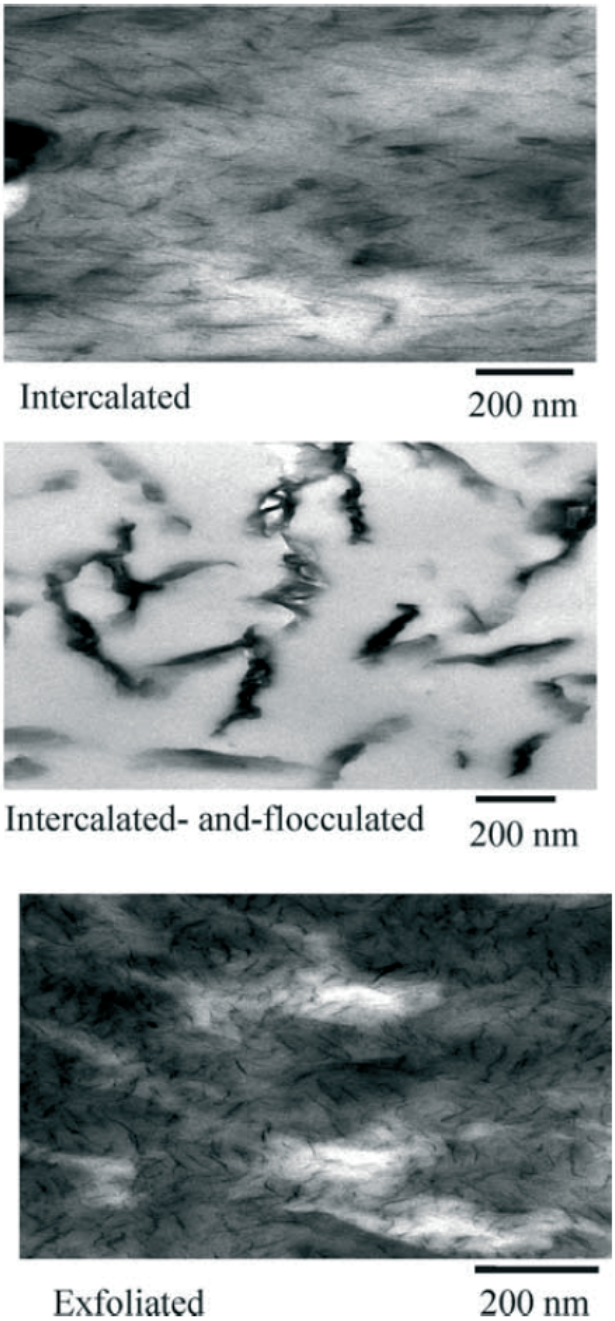


Fig. 3: *Transmission electron microscopy (TEM) images of polymer-silicate nanocomposites. (Reprinted with permission from [13]. Copyright 1997 Elsevier)*

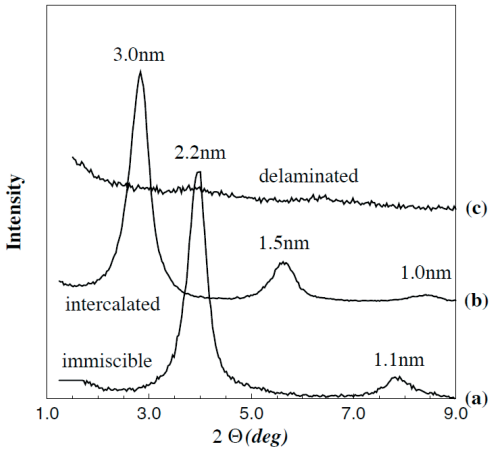


Fig. 4: *Wide-angle X-ray scattering (WAXS) diagrams of polymer-silicate nanocomposites. (Reprinted with permission from Ref. [18]. Copyright 1999 Springer, Berlin Heidelberg New York.)*

More accurately, Fig. 4 shows by recording the position, shape and intensity in scattering diagrams, the intercalated and exfoliated (delaminated) structure can be distinguished very easily. From Chap C1, we know that the corresponding peaks can be considered to be correlation peaks, caused by the distances ξ_{clay} between platelets. We define the momentum transfer $Q \propto \sin\theta$, via the scattering angle θ and remember the reciprocal relationship $\xi_{\text{clay}} = 2\pi/Q$. At first we would like to note that the scattering diagram of an immiscible system is identical to that of the pure clay [18]. The spacing between the platelets is on average 2.2 nm. When a polymer is intercalated the peak typically shifts to smaller θ [13], thus, to greater dimensions. The peak corresponds to 3.0 nm. We note that the additional peaks belong to second and third order reflections, and are not considered further. Furthermore, when the silicate is in its exfoliated (delaminated) state, we do not observe any signature of a peak. Typical distances are outside of the θ values available in typical WAXS experiments. We note that these distances could be revealed by small-angle X-ray scattering (SAXS) or small-angle neutron scattering (SANS) experiments. The respective discussion is outside the scope of this chapter, For details, we refer to the respective literature, e. g. Ref. [26].

This chapter showed that silicates are a good example representing the class of nanocomposites. Therefore, we can exemplify properties of nanocomposites using mixtures based on polymers and silicates. Before discussing experimental results, typical preparation techniques are summarized.

2 Preparation of Nanocomposites

Preparation of nanocomposites can be divided into three major groups: melt intercalation, intercalation of polymer from solution, and *in situ* intercalative polymerization methods. Hereafter these techniques are briefly summarizes. Details can be found in the respective literature, e. g. Ref. [26].

2.1 Melt Intercalation

In many cases, organically modified layered silicates are mechanically dispersed in polymer matrices. Though this is the industrial standard, it could cause chain degradation and the particles are not well dispersed. The item dispersion refers to the spatial distribution of nanoparticles. In order to obtain nanocomposites we need to avoid clustering and to generate isolated nanoparticles. The latter case is referred to as optimum dispersion. Standard industrial mixers do not generate ideally dispersed systems. Therefore, other techniques are required.

2.2 Intercalation of Polymer from Solution

In case of lab-based applications, intercalation from solution seems to lead to better materials, e. g. avoiding chain scission and providing particles that are better dispersed. A solvent is used in which the polymer is soluble and the particles can swell, i. e. the clay platelets can be separated. Typical solvents are water, chloroform, or toluene. At first, both components are put into the solvent and then they are mixed. It is assumed that during mixing the polymer chains intercalate and replace the solvent molecules on the surface of the silicate [13]. After suitably evaporating the solvent we obtain the desired nanocomposite. However, the described procedure has an important disadvantage. In Chap. B1 – Polymers of this textbook, the role of the entropy was highlighted, encompassing its relationship to the number of different conformations due to thermal fluctuations and in different chains. If chains are close to surfaces, the number of available conformations is more limited due to geometrical constraints and confinement effects. Therefore, the polymer chains tend to circumvent this situation and increase the distance to surfaces. This mechanism leads to a 'cavity' close to the nanoparticles that generates a (virtual) attractive force, called depletion force, by which the particles can agglomerate. We assume that the low viscosity in the mixture of polymer, particles and solvent promotes this process.

2.3 *In Situ* Intercalative Polymerization Method

The layered silicate is swollen in the liquid monomer or a monomer solution. Then, appropriate catalysts, e. g. heat, radiation, etc. can initiate polymerization that occurs between the intercalated sheets. Compared to the previous intercalation from polymer solution, the particles are again better dispersed. Currently, this technique seems to mark the optimum.

3 Properties of Nanocomposites

Nanocomposites consisting of a layered silicate and a polymer can show drastically enhanced properties. Hereafter, a few of them are summarized.

3.1 Mechanical Properties

One very famous technique is dynamical mechanical analysis (DMA). It measures the response of the material to an oscillatory deformation. (More details on the method are presented in Chap. C6 of this textbook.) Within the present chapter, we focus on the storage modulus, G', the loss modulus, G'', and $\tan\delta = G''/G'$. G' measures the stored energy, i. e. representing the elastic fraction, and G'' the energy dissipated as heat, representing the viscous fraction. Many authors use $\tan\delta$, because it simply allows observing changes in the molecular mobility, maybe

caused by the glass transition.

Fig. 4 (a) shows a typical example using a mixture of the polymer polylactide (PLA) and MMT, plotting G', G'', and $\tan\delta$ as a function of the temperature. The unfilled material, PLA, is compared with the nanocomposites, with a MMT fraction of 3 (PLACN1), 5 (PLACN2), and 7 wt% (PLACN3). We observe an enhancement of G' by increasing the clay fraction. Inspecting the

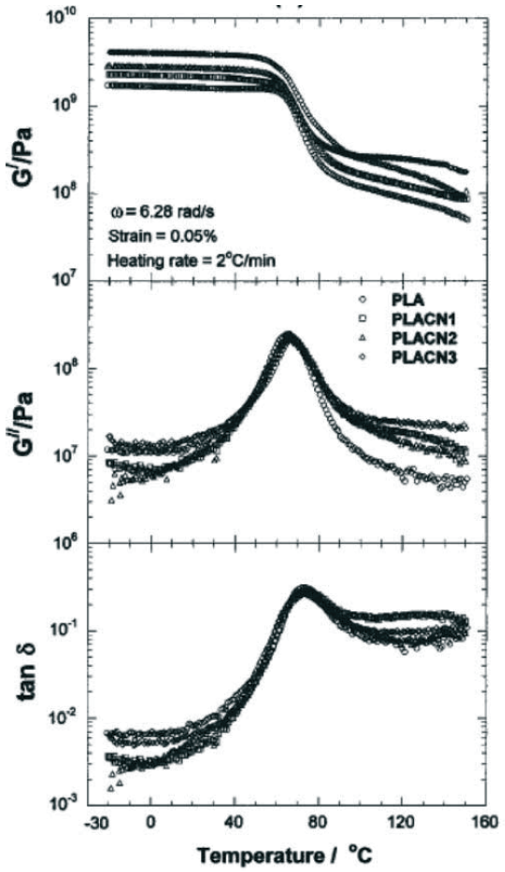


Fig. 5: *Real-, G', and imaginary, G'', part of the complex modulus $G^* = G' + iG''$, and $\tan\delta = G''/G'$ as a function of the temperature for various MMT fractions. (Reprinted with permission from Ref. [19]. Copyright 2002 American Chemical Society.)*

data sets in detail shows a greater enhancement at $T > T_g$ [13]. A mechanical T_g can be defined using the maximum of $\tan\delta$. Fig. 5 demonstrates that the increase of the modulus is a systematic function of the MMT fraction, taking the modulus at $T = 120\,^{\circ}\mathrm{C}$.

Up to now, the enhancement of the modulus, by adding nanoparticles is not fully understood. Currently, it is an up-to-date topic under debate and extensive research is carried out [26].

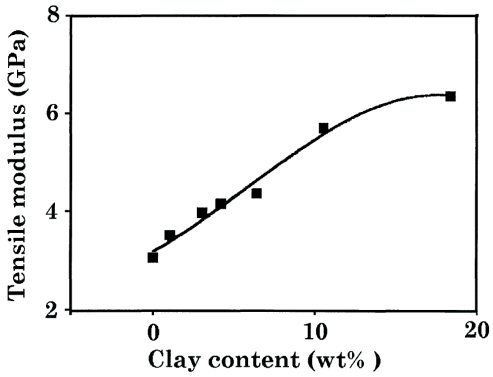


Fig. 6: *Modulus as a function of the MMT fraction at $T = 120\,°C$. (Reproduced from ref. [20] by permission of John Wiley & Sons Ltd)*

3.2 Heat Distortion Temperature

The head distortion temperature (HDT), sometimes referred to as heat deflection temperature, is the temperature at which a polymer sample deforms under a specific load. Due to its importance for product design, engineering and manufacturing of products using thermoplastic components, the testing procedure is standardized in the document ASTM D648 or similarly in ISO 75. The abbreviation ASTM stands for American Society for Testing and Material, which is an international organization that develops technical standards for a wide range of materials. The European equivalent is the International Organization for Standardization (ISO).

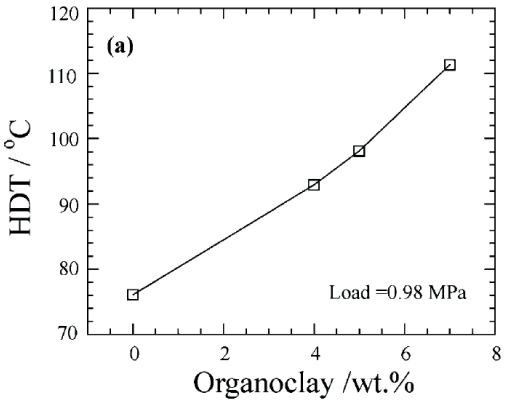


Fig. 7: *Head distortion temperature (HDT) as a function of the silicate fraction. (Reprinted with permission from [21]. Copyright 1997 Elsevier)*

Fig. 6 compares the HDTs of a neat PLA with those of nanocomposites with different MMT fractions. A constant load of 0.98 MPa was applied during the experiment. The result indicates an increase of the HDT from 65 (neat PLA) to $111\,°C$.

It is worth to note that the HDT of nanocomposites is greater than the value in microcomposites [13]. When we recall Fig. 2, we realize that clay platelets possess a high aspect ratio between the thickness and the size of the platelets. The literature revealed that the HDT also depends on that ratio between diameter and thickness of the platelets [14]. Furthermore, the authors conclude that strong hydrogen bonds between polymer chains and solid surfaces of the clay particles determine the HDT.

3.3 Thermal Stability and Fire Retardant Properties

One tool to measure the thermal stability is thermogravimetric analysis (TGA). It simply measures the weight loss of the material during heating. The literature reports a weight loss shifted to higher temperatures when adding nanoparticles. Typically, the degradation occurs at temperatures $50\,^{\circ}$C higher than in the pure polymer [13]. Even with a fraction of 0.1 wt% of clay only, a significant change occurs. Both intercalated and exfoliated materials show an increase of the temperature. It may be explained by introducing barrier properties due to the incorporation of silicate. Details can be found below.

The thermal stability of materials is very important for applications, in particular if a fire occurs. Meanwhile, materials based on the fire retardant properties of nanocomposites, particularly, those based on clays, may help to improve the materials performance. In fact, there are many parameters that can be improved, e. g. the heat release rate (HRR), heat peak HRR, smoke production, etc.

One tool to measure fire retardant properties is the so-called Cone calorimeter. It is a tool that measures the smoke optically and the soot gravimetrically. Details on the method can be found in the respective literature, e. g. Refs. [15, 16]. Fig. 7 presents the enhancement of HRR by the addition of silicates, i. e. a reduced flammability. It shows also a difference for exfoliated and intercalated clays. It is another indication that intercalated materials are more effective than exfoliated polymers. Furthermore, the literature reports that perfect dispersion is essential to show an enhanced HRR [17].

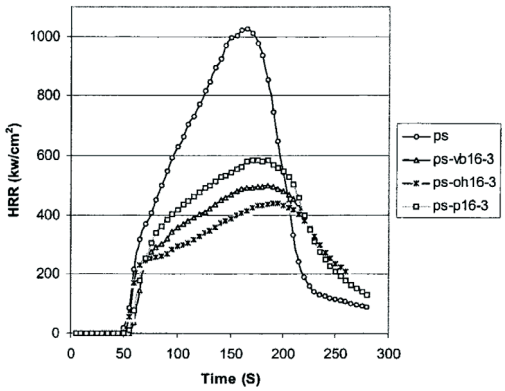


Fig. 8: *Heat release rate (HRR) as a function of the time for various composites with 3 % clay fraction for different clays. (Reprinted with permission from Ref. [22]. Copyright 2001 American Chemical Society.)*

3.4 Gas Barrier Properties

In Section 1 of the present Chapter, we became aware of the requirement to control the gas barrier properties in materials, e. g. for having a perfect 90 minutes soccer game. To study this water vapor permeability tests of nanocomposites films are suitable (ASTM E96-95) [23]. For that purpose dry nanocomposite films very used to seal the open flange of bottles and put it under certain conditions, such as a constant temperature and humidity. From the weight of the bottle the water transmission rate can be calculated and from that the permeability.

Fig. 9 shows the relative gas permeability P_c/P_0, i. e. the gas permeability P_c divided by the gas permeability of the pure material P_0. The symbols show a significant reduction of the gas permeability with increasing the silicate fraction. The theoretical lines assume well dispersed nanoparticles and evaluate the predictions using different aspect ratios $\alpha = 300$, and 1000 .

From such experiments, it became clear that clays introduce barrier properties [23]. Furthermore, the barrier properties seem to be different when the clays form intercalated or exfoliated structures [24]. This observation is important, because it indicates that besides the trivial change due to the increase of the particle fraction the structure or eventually the orientation plays an important role.

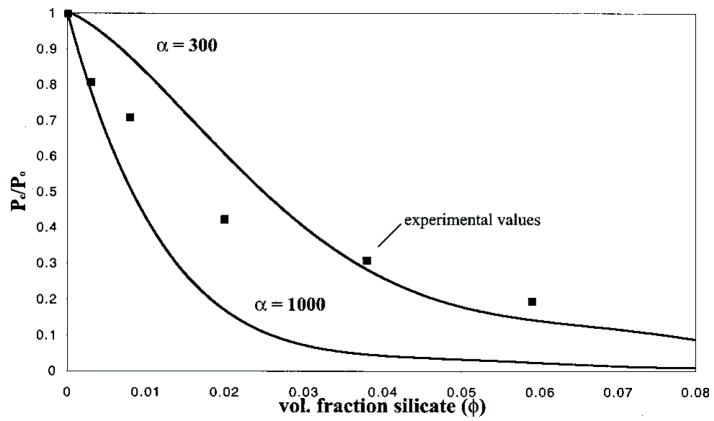


Fig. 9: *Relative water permeability as a function of the silicate volume fraction of silicate. The symbols are experimental values and the full lines are calculated. (Reprinted with permission from Ref. [23]. Copyright 2001 American Chemical Society.)*

In order to relate the morphology with the gas permeability, simulations are very useful [24]. Fig. 10 shows the influence of the orientation of platelets, as described by the order parameter S. The simulations reveal that the relative permeability depends on the order parameter S. Additionally the dependence on the diameters L of the platelets is visualized. The thickness $W = 1$ is kept constant. Furthermore, the literature shows that the intercalated structure is less desired compared to the exfoliated structure, when low permeabilities required [24].

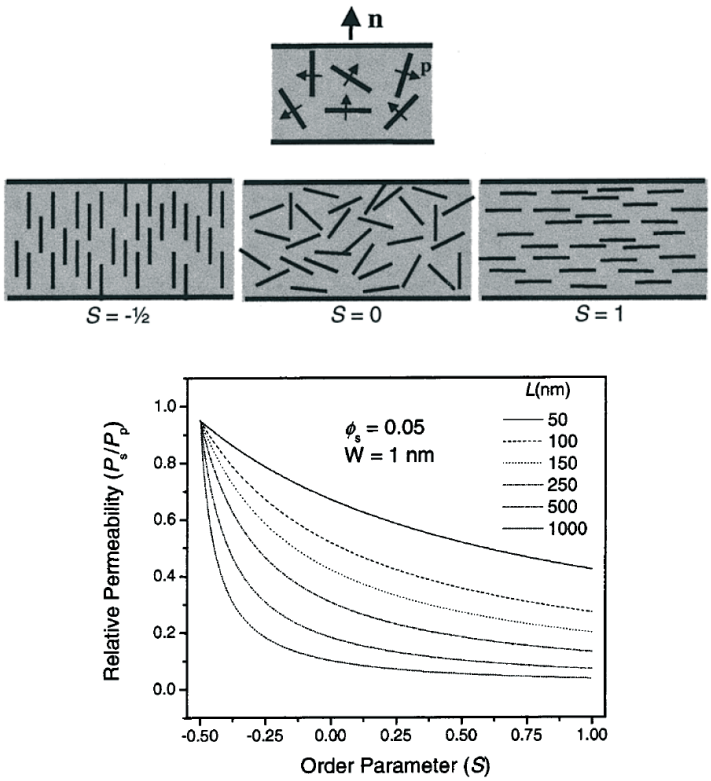


Fig. 10: *Influence of the platelet orientation described by the order parameter S on the relative permeability for various diameters L of the platelets. (Reprinted with permission from Ref. [24]. Copyright 2001 American Chemical Society.)*

3.5 Optical Transparency

Obviously, the properties presented depend on the nanoscale dispersion of the clays into the material. Besides these enhancements, going from micro to nanoparticles changes the optical properties from opaque to transparent. For example, due to diameters much smaller than the wavelength of visible light, light is not scattered and therefore true nanocomposites are optically transparent. This speculation can be simply tested by UV/vis transmission spectra, recording the transmittance as a function of the wavelength. The experimental data show no or almost no changes within the region of visible light ($400 - 700\,\mathrm{nm}$). It also shows that for smaller wavelengths the absorption is affected, because here the clay size starts to play an important role.

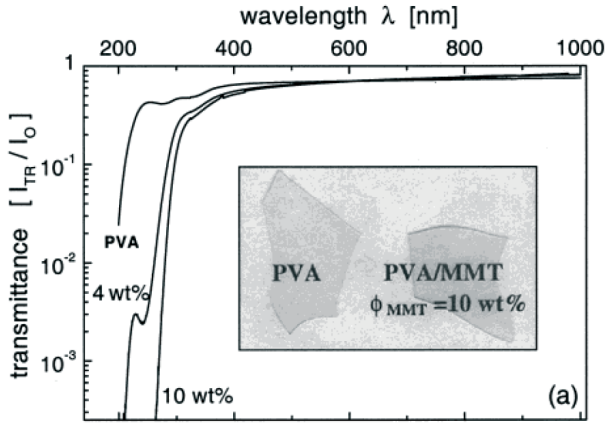


Fig. 11: *UV/vis transmission spectra as a function of the wavelength, for different silicate fractions. (Reprinted with permission from Ref. [25]. Copyright 2000 American Chemical Society.)*

3.6 Conclusions

The successful stories of nanocomposites could be continued, e. g. possible changes in ionic conductivity, biodegradability, crystallization, etc. Most likely, these few examples, will be accompanied by many novel and unprecedented properties that we can not imagine today. Therefore, the research on nanocomposites will continue and contribute to novel and important applications in the future.

References

[1] C. Sanchez, B. Julian, P. Belleville, M. Popall, J. Mater. Chem. *15*, 3559 (2005)

[2] R. Wengenmayer, Max-Planck-Forschung *3*, 52 (2001)

[3] U. Hänsler, Max-Planck-Forschung *4*, 38 (2002)

[4] P. Gómez-Romero, C. Sanchez, New J. Chem. *29*, 57 (2005)

[5] E. Fois, A. Gamba, A. Tilocca, Microporous and Mesoporous Mater. *57*, 263 (2003)

[6] D. Reinen, P. Kohl, C. Müller, Z. Anorg. Allg. Chem. *630*, 97 (2004)

[7] http://lanxess.com/en/rubber-for-soccer-balls/ (15/12/2014)

[8] H. J. Gao, B. H. Ji, I. L. Jäger, E. Arzt, P. Fratzl, Proceedings Of The National Academy Of Sciences Of The United States Of America *100*, 5597 (2003)

[9] K. Nusser, T. Mosbauer, G. J. Schneider, K. Brandt, G. Weidemann, J. Goebbels, H. Riesemeier und D. Göritz, J. Non-Cryst. Solids *358*, 557 (2012)

[10] H. D. Luginsland, *A review on the chemistry and the reinforcement of the silica-silane filler system for rubber applications* (Shaker Verlag, Aachen, 2002)

[11] G. M. Kim, *Verstärkungsmechanismen auf Makro-, Mikro und Nano-Längenskalen in heterogenen Polymerwerkstoffen* (Habilitationsschrift, Martin-Luther-Universität Halle-Wittenberg, 2007)

[12] X. Zhu, J. Phys. Condens. Matter *15*, L253 (2003)

[13] S. Sinha Ray, M. Okamoto, Prog. Polym. Sci. *28*, 1539 (2003)

[14] Y. Kojima, A. Usuki, M. Kawasumi, A. Okada, Y. Fukushima, T. Kurauchi, O. Kamigaito, J. Mater. Res. *8*, 1185 (1993)

[15] C. Huggett, Fire and Materials *4*, 61 (1980)

[16] W. H. Twilley, V. Babrauskas, *User's Guide for the Cone Calorimeter* (NBS Special Publication 745, U. S. Department of Commerce, National Bureau of Standards, Washington, 1988)

[17] J. W. Gilman, C. L. Jackson, A. B. Morgan, R. Harris Jr., E. Manias, E. P. Giannelis, M. Wuthenow, D. Hilton, S. H. Phillips, Chem. Mater. *12*, 1866 (2000)

[18] E. P. Giannelis, R. Krishnamoorti, E. Manias, Adv. Polym. Sci. *138*, 107 (1999)

[19] S. Sinha Ray, P. Maiti, M. Okamoto, K. Yamada, K. Ueda, Macromolecules *35*, 3104 (2002)

[20] L. Lium Z. Qi, X. Zhu, J. Appl. Poly. Sci. *711*, 1133 (1999)

[21] S. Sinha Ray, K. Yamada, M. Okamotoa, K. Ueda, Polymer *44*, 857 (2003)

[22] J. Zhu, A. B. Morgan, F. J. Lamelas, C. A. Wilkie, Chem. Mater. *13*, 3774 (2001)

[23] R. Xu, E. Manias, A. J. Snyder, J. Runt, Macromolecules *34*, 337 (2001)

[24] R. K. Bharadwaj, Macromolecules *34*, 9189 (2001)

[25] K. E. Strawhecker, E. Manias, Chem. Mater. *12*, 2943 (2000)

[26] S. Sinha Ray, *Clay-Containing Polymer Nanocomposites: From Fundamentals to Real Applications* (Elsevier, Amsterdam, 2013)

D 4 Microemulsions

H. Frielinghaus
Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Concepts for Aqueous Surfactant Solutions **2**

3 Concepts for Microemulsions **6**

4 Phase Diagrams of Microemulsions **13**

5 Polymer Boosting Effect **16**

6 Microemulsions Near Planar Walls **22**

7 Summary **25**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Microemulsions are both: ideal systems for academic research and technical solutions for cleaning processes. Classically, the emulsion is a related system that reaches kinetic stability by an amphiphile that mediates between water and oily substances. For many applications the time-limited stability is sufficient, such as food products, cosmetics, and paints. Either the degradation is wanted, or simple stirring can restore the initial state. Microemulsions are thermodynamically stable. This facilitates the formulation, because all times the same state is obtained. For academic research this is highly important to make experiments reproducible and highly precise. For industrial products the stability is also desirable. The introduced paint remover often stays at home for longer periods before it is used another time. Each time, the product performs equally good.

Microemulsions consist of three elementary substances: water, oil and a surfactant. The surfactant mediates between the two immiscible components and allows for a stable mixture. This is directly seen visually, by obtaining a transparent liquid. Microscopically, there are still domains of water and oil, and the surfactant forms a thin film between the two domains. The structures are in the nanometer range, and this is why scattering experiments are the ideal tools for resolving the structure.

Amphiphilic polymers have been found to be ideal additives to microemulsions. The efficiency of the surfactant is dramatically increased, and so large amounts of surfactant can be saved. This makes applications cost effective, because surfactants are usually expensive. Furthermore, the cleaning abilities might be increased as well.

The other advantage of microemulsions is the structural size that is highly reproducible by formulating the complex fluid specifically. This is needed, if chemical reactions take place in one of the aqueous of oily domains. Nano-reactors host reactions for nano-scale products that have determined size and low polydispersity [1]. Apart from that, specific foams can be obtained with well defined and small bubbles. So, very good thermal insulators can be obtained at low cost [2]. Also for water-fuel mixtures that inhibit large amounts of soot and oxides of nitrogen the droplet size of the water droplets matters: In the combustion, the fuel sprays more homogenously for smaller droplets [3]. Many of these applications wait for inexpensive solutions where the structure needs to be resolved using scattering experiments.

In this manuscript there are concepts presented first: for aqueous surfactant systems and for microemulsions. The basic terms are introduced and related by the underlying concepts. The three most often used phase diagrams are presented and their meaning to science and applications is discussed. The most important additive, the amphiphilic polymer, is discussed in detail: conceptually and with respect to applications. Apart from that, the studies of microemulsions adjacent to planar walls open new perspectives of tailoring microemulsions for specific applications.

2 Concepts for Aqueous Surfactant Solutions

Surfactants can be divided into two major classes: Ionic surfactants possess a ionic head group with a counterion while non-ionic surfactants have no charges. In the first case, the ionic head group is soluble in water, and the counterion dissociates. One for research important surfactant is the sodium dodecyl sulfate (SDS, see Fig. 1). The SDS is an anionic surfactant because the sulfate head group is an anion. The tail of the SDS molecule is a hydrocarbon, which is typical

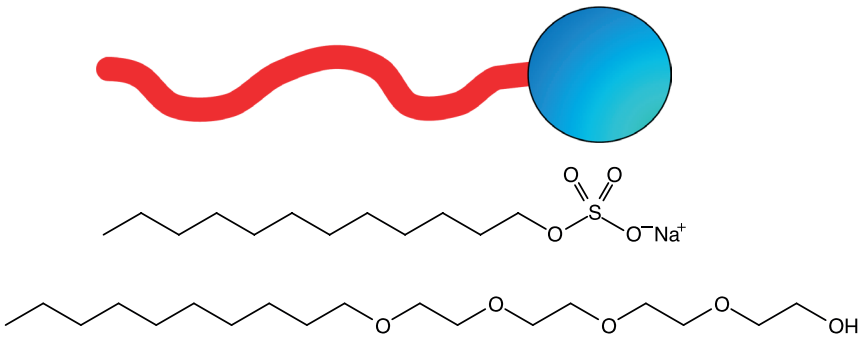


Fig. 1: *Top: Conceptual drawing of a surfactant molecule. The hydrophilic head is blue while the hydrophobic tail is red. Middle: Chemical structure of sodium dodecyl sulfate (SDS). The head and tail groups are just below the conceptual drawing. Bottom: Chemical structure of tetraethyleneglycolmonodecylether ($C_{10}E_4$). The head group is four glycol groups long.*

for most of the surfactants. A representative for the non-ionic surfactants is tetraethyleneglycolmonodecylether ($C_{10}E_4$, the indices count the carbon atoms and ethylene oxide groups, see Fig. 1). The head group is not charged; but the oxygen atoms along the head group give rise to hydrogen bonding, which is favorable for the water solubility. For this type of surfactant the molecule ends can vary in length, but also in chemical structure. For instance the hydrophobic tails can possess different amounts of saturated carbon-carbon bonds. This is important for lipids, which are natural ionic surfactants forming cell membranes. Lipids often have two hydrophobic tails. The number of double-bonds in the tails determines the thermodynamic state of the membrane. Many unsaturated tails give rise to crystalline order of the hydrophobic tails [4]. Apart from the tails, the head groups may possess two oppositely charged groups; then the surfactant is called amphoteric. The whole concept of hydrophilic and hydrophobic can be extended by a third type of philicity: The polymer Teflon (fluorinated carbon chains) is known to be neither water soluble nor oil soluble. If fluorinated carbon chains are used as hydrophobic tails a new class of surfactants is obtained[1]. Throughout this manuscript we limit ourselves to the simple twofold concept of hydrophilicity and lipophilicity. The interested reader may find further information about fluorinated surfactants in the literature [5].

We now consider aqueous solutions of a single surfactant type. It is known that at very low concentrations the surfactant molecules are dissolved independently. The reason for this behavior is the entropy, which favors dissociated molecules. But because the hydrophobic tail causes some enthalpic violation, at the critical micelle concentration (CMC) the surfactant molecules associate and form small spherical micelles. The hydrophobic tails are in the center and the hydrophilic heads surround the micelle. The hydrophobic neighborhood of the hydrocarbon chains can be monitored by NMR [6] and so very precise values for the CMC can be given. The effect of the CMC is a volume effect and is thus determined for large volumes. At the surface, the surfactant molecules can also be found. These studies focus on Langmuir-Blodgett films for instance, but this topic will lead too far.

The formation of micelles corresponds to the condensation of gases to small droplets with one difference: The micellar dimensions are determined by the molecular size of the surfac-

[1] Fluorinated surfactants allow for CO_2 to be used as hydrophobic component in microemulsions for instance.

Table 1: *The different micellar structures predicted on the basis of the packing parameter.*

P	molecule geometry	micelle structure	symmetry
$< \frac{1}{3}$	cone	sphere	point-like
$\frac{1}{3}$ to $\frac{1}{2}$	wedge	cylinder	cylinder
$\frac{1}{2}$ to 1	wedge	vesicle (double layer)	point-like
1	cylinder	planar double layer	plane-like

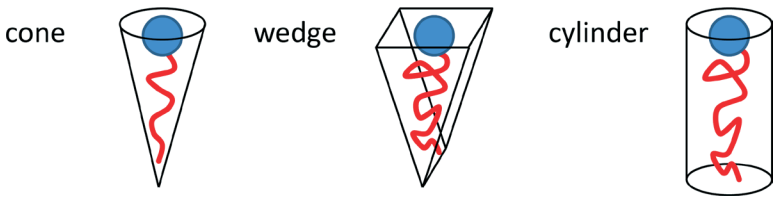


Fig. 2: *Molecule geometries for different packing parameters.*

tant while condensed droplets can take any size. The example of micelles is typical for self-assembling systems.

The next question focuses on the state or structure of the micelles in solution. Different structures can be classified and shall be explained on the basis of a simple model, which mainly focuses on ionic surfactants. The parameter of interest is the packing parameter [7], which is defined as follows:

$$P = \frac{v}{a \cdot l} \tag{1}$$

In this equation the parameter v is the volume of the whole molecule, a is the area of the head group, and l is the length of the chain. This packing parameter can vary between values below $\frac{1}{3}$ and 1 (see Table 1). For values below $\frac{1}{3}$ the micelles are spherical, then for P up to $\frac{1}{2}$ the micelles are elongated cylinders. For P up to 1 the micelles form closed double layers, i.e. spherical hollow membranes; they are called vesicles. For $P = 1$ the membranes become planar. The formed structures have a high degree of symmetry. Only fluctuations might destroy the high degree of symmetry. So for instance very long cylindrical micelles start to bend and a worm-like micelle is formed [8]. For the purpose of this lecture we restrict the considerations of P to a maximum of 1.

An experimental phase diagram is depicted in Fig. 3. We first restrict ourselves to the temperature of 20°C (see Fig. 3). The CMC is found at concentrations of around 0.005%. At higher concentrations up to ca. 1% spherical micelles are found. In between 1 and 10% the micelles become cylindrical. Going up in temperature now, their length grows until the phase boundary at ∼33°C is reached. A clear line between spherical and cylindrical micelles is not given in the phase diagram because the phase transition smears out, and usually a coexistence between the two morphologies is found. The long micelles are usually wormlike because of the fluctuations. At higher concentrations the worms can even form networks. All these effects take place in the one-phase region (1Φ or L_1). The temperature has an effect on the micelle shape because at lower temperatures water penetrates the head group of the non-ionic surfactant.

So far we did not consider the case, that the micelles can be reversed. At low water concentra-

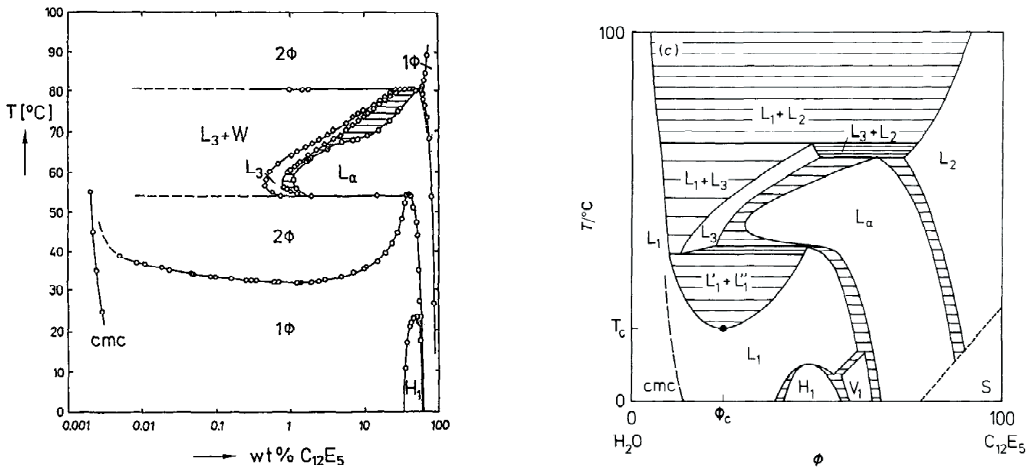


Fig. 3: *An experimental phase diagram of the non-ionic surfactant $C_{12}E_5$ in water [9, 10]. Following a horizontal line at ca. 20°C from low to higher concentrations one finds the CMC at around 0.005% and the one-phase (1Φ) region. Below ∼1% the micelles are spherical and uncorrelated. Between 1 and 10% the micelles become cylindrical. Their length increases with temperature until the phase boundary at around 33°C is reached. Interestingly, more phases are found than predicted by the simple packing parameter approach. 2-phase coexistence is indicated by 2Φ. The hexagonal phase is indicated by H_1 and the lamellar phase by L_α. The L_3-phase is the sponge phase. To the right the same diagram is shown on linear scale and more schematically [9]. The abbreviations for the different phases are discussed in the text below (see also Table 2).*

tions (or high surfactant concentrations) the water and hydrophilic heads form a closed volume surrounded by the hydrophobic parts. The corresponding region is indicated by L_2 or 1Φ. The interesting case of the L_3-phase is found at slightly higher temperatures. Then the membranes fill the whole volume with a sponge like structure. The membranes are strongly fluctuating. The more planar membranes are found in the L_α-phase at relatively high concentrations. These lamellae are relatively well ordered due to steric interactions. Steric interactions are typical for non-ionic surfactants, which do not possess a Coulomb interaction. For ionic surfactants the lamellar phase is formed at lower concentrations due to the strong interaction. Astonishingly, more ordered phases appear. The H_1-phase contains cylindrical micelles (as the L_1-phase), but the cylinders are ordered on a hexagonal lattice. Again, the steric repulsion is sufficient to order the micelles in a liquid crystalline state. The V_1-phase has a cubic unit cell, while the hydrophilic and hydrophobic domains are continuous in the whole volume. This phase is also called 'plumbers nightmare' because the high viscosity might result in plugging of tubes. The exact structure will be discussed in the text below. We see that liquid crystalline phases (L_α, H_1, V_1) exist with liquid crystalline order. These phases are also termed lyotropic because the addition of water is responsible for the formation. So this term arises from the viewpoint of the solid phase S, which means mainly a pure surfactant with small impurities of water.

In Fig. 3 there are also two-phase regions marked. In this region two phases coexist, so either the sample gets turbid because of many small domains or, after a long time, the sample forms a meniscus between the two clear phases. The horizontal lines indicate the corresponding co-

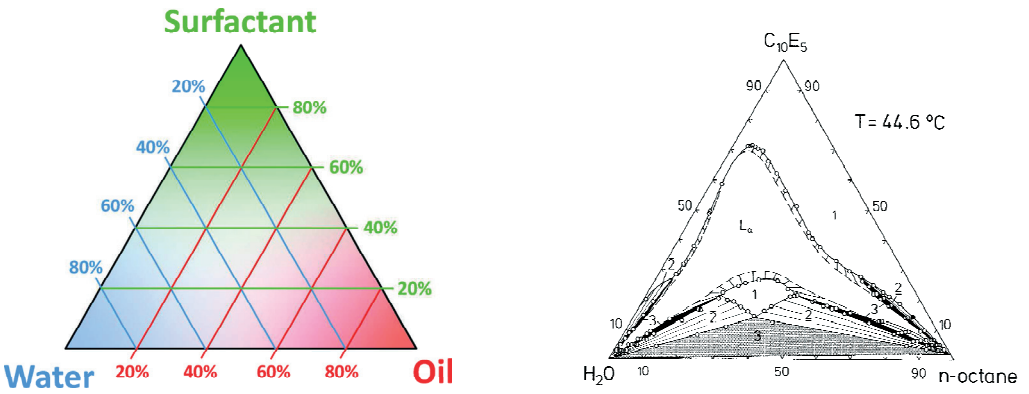


Fig. 4: *Left: A scheme how to read off the compositions from the Gibbs Phase Triangle. The axis to the left of the corresponding component in the corner shows the scale of the fraction. The colored lines indicate constant composition. The whole diagram is completely symmetric. Right: A rather simple phase diagram for the system: H_2O, n-octane, $C_{10}E_4$ [11]. The numbers inside indicate the number of phases, and L_α indicates the lamellar phase.*

existing phases, so from a given overall concentration one follows the tie-lines to the right and left, and reads off the properties of the coexisting phases. The lines are horizontal, because the vertical axis is the temperature. For more complicated phase diagrams (we shall see later) the tie lines can be tilted. The example $L_3 + W$ or $L_1 + L_3$ indicates a coexistence of the sponge phase L_3 with a highly water rich phase. The example $L_1' + L_2''$ or 2Φ indicates two coexisting micellar phases. One of them contains spherical and the other one cylindrical or wormlike micelles.

In summary, for the aqueous surfactant systems, the following points shall be clear: The CMC separates the highly diluted from the diluted region. The entropy favors unimeric surfactant molecules while the enthalpy favors micelles. The concept of packing explains the micellar shapes. These shapes have a high degree of symmetry, because each surfactant molecule is identical. The shapes range from spherical over cylindrical to lamellar. At higher concentrations, the interactions between the micelles lead to lyotropic (or liquid crystalline) phases. There exist phases with the same micellar shapes of the diluted region, but also ordered phases with new unit cells (see V_1). The interactions are sterically repulsive for non-ionic surfactants and Coulomb-like for ionic surfactants. Theoretical concepts of the interactions will be given in the following chapter. The parameter temperature comes into play here because enthalpic and entropic contributions are weighted differently.

3 Concepts for Microemulsions

So far we have been focusing on two-phase systems. For cleaning processes the uptake of oil is an important issue. Then microemulsions will be formed. It is known that on a microscopic level there are domains of (nearly) pure water and oil, and the surfactant is at the interface. In this sense, the surfactant mediates between the hydrophilic and hydrophobic components, which leads to macroscopically homogenous fluids. Thus, microemulsions are mostly optically clear, which is one criterion for phase diagram measurements.

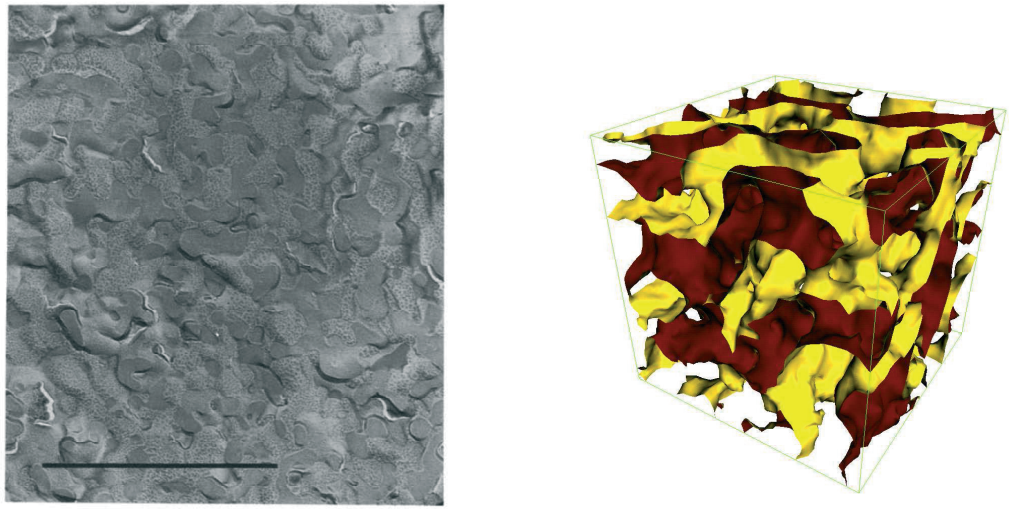


Fig. 5: *Left: A transmission electron micrograph of the microemulsion containing water, octane, and $C_{12}E_5$. The surfactant content was 7 wt% (see Ref. [12]). The indicated bar shows a scale of 1μm. Right: A real space picture of the bicontinuous microemulsion according to computer simulations [13]. Actually the surfactant film is shown with the surface color being red for oil facing surface and yellow for water facing surface.*

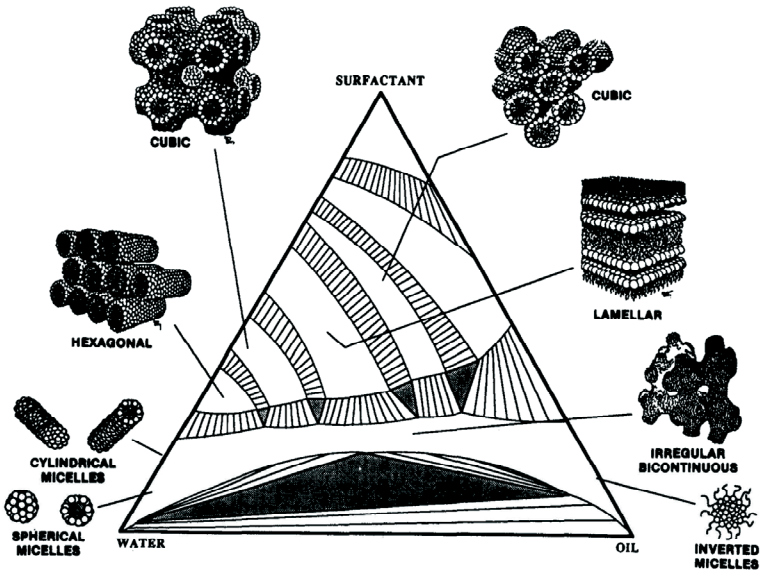


Fig. 6: *A more schematic phase triangle with most typical lyotropic phases [14]. At the bottom there is the three-phase coexistence region (black) surrounded by two-phase coexistence regions. Above there is the large L_1-region with droplets, cylinders and the bicontinuous phase. In the upper half there are many lyotropic phases such as the hexagonal H_1, the cubic V_1, the lamellar L_α, and the fcc or bcc cubic I_1 phase.*

A rather simple phase diagram is shown in Fig. 4. There are now one-phase, two-phase, and three-phase coexistence regions, since now three components are used. The most interesting region is the small one-phase region almost in the center of the phase triangle. Here the bicontinuous microemulsion is found. The components oil and water both form a sponge-like structure, i.e. each of the sponge hosts the other one. In this sense the phase is *bi*-continuous (see also Fig. 5). This can for instance be proven by conductivity measurements. The grey three phase coexistence region in the bottom of the triangle indicates a coexistence of a water-rich, an oil-rich, and a bicontinuous phase. Here the concept of tie-lines breaks down. Contrarily, all two-phase regions are filled with tie-lines, which are tilted now. At very low surfactant concentrations the droplet phase is found. This phase is nearly invisible on the current scale. A more schematic phase diagram with lyotropic phases is depicted in Fig. 6. One important point is the large L_1 phase where droplets, cylinders, and the bicontinuous phase are included. Entropic contributions destroy clear phase transitions between the distinct structures, and so coexistence is possible. On the bottom right the reversed micelles are found. Another point of this scheme is the indication the most important lyotropic phases. Their real space pictures are indicated around the phase triangle. The structures show closed micelles with the hydrophobic component inside. In principle, the reversed micelles are possible as well. After we have seen that different authors use different abbreviations for the same phases a list of all possible abbreviations should be given. A first attempt was made by Tiddy [4] that we now extend for our own purposes (see Table 2).

While the theoretical concept for aqueous surfactant systems describes the micelles as bulky objects the most widely accepted concept for microemulsions bases on the theory of Helfrich [16]. Here it is assumed that the surfactant forms a membrane and the free energy of the overall system is dominated by the elastic properties of the membrane. The free energy reads then:

$$F = \int dS \left(\gamma + \frac{1}{2}\kappa(c_1 + c_2 - 2c_0)^2 + \bar{\kappa} c_1 c_2 \right) \tag{2}$$

The first addend describes the surface tension of the membrane. In principle the surfactant might vary the formed surface by different tilt angles (the molecules are not oriented perpendicular) or by crystallization of the hydrophobic tails. For our purposes we assume a liquid membrane, and neglect variations of the overall surface. The next summand is a product of the bending rigidity κ and the deviation of the mean curvature $\frac{1}{2}(c_1 + c_2)$ from the equilibrium curvature c_0. The curvature arises from a tangential construction at a given membrane point (see Fig. 7). Namely, two perpendicular circles describe the tangent. Their reciprocal radius is the curvature, i.e. $c_i = R_i^{-1}$. A positive curvature means a curvature towards the oil domain. The middle summand is sensitive to deviations of the mean curvature from the equilibrium curvature. The last addend is a product of the saddle splay modulus $\bar{\kappa}$ and the Gaussian curvature $c_1 c_2$. A saddle shape for instance has a negative Gaussian curvature, while for a sphere the Gaussian curvature is positive, i.e. $c_1 c_2 = R^{-2}$. One finds typical values of $\kappa \approx 1..10 \mathrm{k}_B\mathrm{T}$ and $\bar{\kappa} \approx -\kappa$ for soft to rigid membranes. On this basis, predictions for the phase behavior can be made as we will see below.

The first problem to tackle is the L_1 phase. As we have seen, there exist spherical and cylindrical micelles. The lamellar phase L_α will be taken into account as well. The problem was treated by Safran [17] for the first time by comparing the free energies for the three different cases. Since the bodies were assumed to be ideally shaped the calculations were kept quite simple, i.e. all surface integrals are carried out for constant curvatures. He found that the three different shapes are separated by distinct phases. The same problem was described by Blokhuis [18]

Table 2: *A survey about symbols for the different phases in aqueous surfactant systems and microemulsions. A first attempt was introduced by Tiddy [4]. Especially more exotic examples are given there. The second column gives symbols for polymeric systems [15] for curiosity.*

symbol used here	symbol (polymers)	alternative symbols	explanation
1, 1Φ		L_1 or L_2	micelles & fluctuating bicontinuous phase
2, 2Φ		$L_1' + L_1''$, ...	2 coexisting phases (need to be specified)
$\bar{2}$, $\underline{2}$			2 coexisting phases at high/low temperatures
3, 3Φ		$L_1 + L_2 + L_3$, ...	3 coexisting phases (need to be specified)
L_1	M_1		micelles, hydrophopic part inside
L_2	M_2		reversed micelles, hydrophilic part inside
L_3			bicontinuous phase
L_α	L	D, G	lamellar phase, ordered
H_1	H_1	E, H_I, M_1	hexagonal phase, ordered
H_2	H_2	F, H_{II}, M_2	reversed hexagonal phase, ordered
I_1	C_1	Q_I, S_{1c}	cubic phase$_{\mathrm{fcc,bcc}}$ with spherical micelles
I_2	C_2	Q_{II}	cubic phase$_{\mathrm{fcc,bcc}}$ with rev. spher. micelles
V_1		I_1', Q_I	cubic phase with bicontinuous structure
V_2		I_2', Q_{II}	cubic phase with rev. bicont. structure
	G_1		cubic gyroid phase
	G_2		cubic gyroid phase, reversed

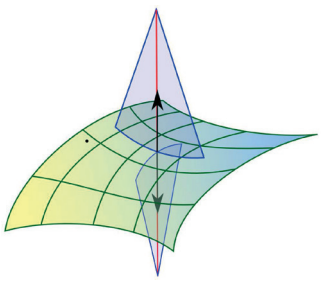


Fig. 7: *An example of a surface with the two principal radii indicated. This construction can be done for any point of the surface.*

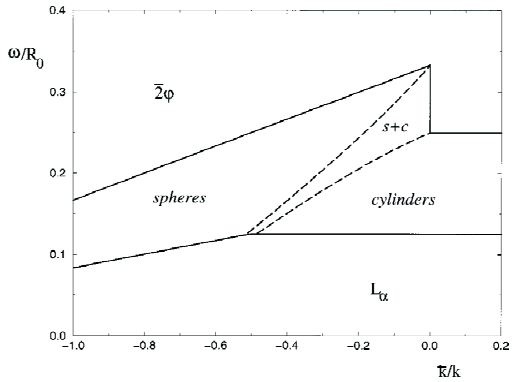


Fig. 8: *Microemulsion phase diagram [18]. The parameter ω/R_0 is given by the ratio of the total volume and the membrane surface and the equilibrium curvature, i.e. $\omega/R_0 = V_{\mathrm{tot}}/S_{\mathrm{tot}} \cdot c_0$. The x-axis shows the ratio of the two moduli $\bar{\kappa}/\kappa$.*

in a slightly extended way: Emulsification failure and coexistence of the two micelle types was taken into account. The results are shown in Fig. 8. On the y-axis the dimensionless ω/R_0 is used. It bases on the ratio of the toal volume and the total membrane area, i.e. $\omega = V_{\mathrm{tot}}/S_{\mathrm{tot}}$, and the equilibrium radius $R_0 = c_0^{-1}$. The x-axis is spanned by the ratio of the two moduli $\bar{\kappa}/\kappa$. From small to large surfactant concentrations one passes from the two phase coexistence ($\bar{2}\varphi$) over the micellar shapes (spheres/cylinders) to the lamellar phase. The choice of the micellar shape is driven by the ratio of $\bar{\kappa}/\kappa$. This means if $\bar{\kappa}$ is strongly negative the spherical micelles are favored. Cylinders (with no Gaussian curvature) are favored from slightly negative to positive $\bar{\kappa}$-values. The theory of Blockhuis was extended to include the translational entropy and polydispersity of the geometrical dimensions. This example shows nicely that entropy smears out the transition between spherical and cylindrical micelles. At this stage of the theory the micelles do not interact.

The interactions to be considered are either sterically repulsive or long ranged Coulomb interactions. The treatment usually involves approximations in different ways. We will introduce two different methods in this manuscript. Schwarz and Gompper [19, 20] considered different minimal surfaces on a cubic lattice. The principal structures are known already (see Fig. 9). For such surfaces the elastic energy as given in equation 2 is minimal with respect to the boundary conditions. In principle, such surfaces were also used as decorative architecture, for instance for the Olympic stadium in Munich. They can be understood as soap bubbles, which form the shape due to the surrounding (boundary condition) and the surface tension. The different minimal surface energies of the cubic symmetry need to be calculated and compared for the different structures. Interestingly, thermal fluctuations can approximately be taken into account. The additional free energy term reads then:

$$F_{\mathrm{steric}} \propto c_0^{-2} \left(\frac{k_B T}{\kappa}\right)^2 \frac{\phi_{\mathrm{oil}}^3}{(1-\phi_{\mathrm{oil}})^2} \tag{3}$$

This energy depends on the bending rigidity κ and the oil volume fraction ϕ_{oil}. For large κ the fluctuations are suppressed, and the additional free energy becomes small. Small equilibrium curvatures c_0 and large oil fractions ϕ_{oil} may make the steric term large. The result of this calculations is that the cubic structures *G*, *D*, and *P* are favored with respect to the other cubic structures taken into account (see Fig. 9).

The example of eq. 3 for steric repulsion of membranes without charges was first derived for planar membranes [21]. The principal interaction is called Helfrich interaction and takes the following expression:

$$V_{\mathrm{steric}}(r) = \frac{\beta (k_B T)^2}{\kappa}\, r^{-4} \tag{4}$$

It bases on the fluctuations of the neighboring membranes that interact through steric repulsion. Following the arguments of Helfrich, β takes the value $9\pi^2/128 \approx 0.68$, but from Monte Carlo simulations a lower value of 0.32 was found. Apart from that, for charged planar membranes the following interaction potential is obtained:

$$V_{\mathrm{el}}(r) = \frac{\pi k_B T}{4\lambda_{\mathrm{B}}}\, r^{-3} \tag{5}$$

with the Bjerrum length being $\lambda_{\mathrm{B}} = e^2/(\epsilon k_B T)$, which takes ca. 7Å in the case of water. This potential arises from a pure electrostatic approach which neglects the thermal fluctuations of

the membrane, and so in reality, a superposition of both potentials would apply for charged membranes.

Another approach for describing phase diagrams of microemulsions bases on a Landau expansion. For this purpose the order parameter Φ needs to be defined. Inside the whole sample the function $\Phi(r)$ takes values between -1 and $+1$. The extreme cases indicate pure oil and pure water domains. Since the function is continuous intermediate values exist in between. These values are usually interpreted as the presence of surfactant. Pure surfactant would mean $\Phi = 0$ while intermediate values are interpreted as mixtures of oil or water with the surfactant. This modeling is contradicting in two aspects: First, the domains of oil and water have usually sharp boundaries and the order parameter would be discontinuous. Second, the nearly incompressible fluid would actually need two order parameters to describe the physics completely. For simplicity reasons and due to its success, the simple model is still often used in the literature [22]. Generally, the Landau approach is very successful in describing fluctuations and phase transitions in solid state physics, soft matter physics and more remote fields. The free energy functional was kept dimensionless in reference [22], and it reads:

$$F[\Phi] = \int dV \left(-\frac{\chi}{2}\Phi^2 + \frac{1-\Phi}{2}\ln\frac{1-\Phi}{2} + \frac{1+\Phi}{2}\ln\frac{1+\Phi}{2} - \frac{1}{2}(\nabla\Phi)^2 + \frac{1}{2}(\nabla^2\Phi)^2 - \mu\Phi \right) \tag{6}$$

The first addend is a simplified treatment of interactions on the basis of a point like interaction with the interaction parameter χ. It is fully correct for steric repulsions, and also for polymeric systems with only next neighbor interactions. Coulomb interactions would need a distance dependent interaction. The next two terms arise from the translational entropy of the oil and water domains (The size of the molecules is assumed to be identical). Actually, these two terms do not follow strictly the concept of a Landau approach, because then only a Taylor expansion of this expression would appear. The next two terms arise from the functional expansion of the order parameter. Odd terms do not appear due to the high symmetry of the system (usually assumed; for instance a gradient term could describe gravity effects). The gradient term describes the low surface tension of the system. The negative sign means that certain surfaces between domains are favored (especially on large length scales). The next order correction sets a limit to these surfaces (at small length scales the homogenous state is favored). The last term describes the chemical potential describing the conjugated field [23]. In this way the phase diagram can be displayed as a function of the mean order parameter or the conjugated field. The direct prediction is the existence of lamellar L_α and hexagonal H_1 and H_2 fields (see Fig. 10). For such a phase diagram either different ordered fields $\Phi(r)$ with sinusodial oscillations are assumed analytically and their free energy is compared on the basis of the integral (eq. 6). A better approach is obtained by computer calculations of $\Phi(r)$ on a lattice. The computer can take higher order oscillations into account more easily. Furthermore, a computer can simulate thermal fluctuations relatively straight forward, while analytically the effort is often relatively high, especially for the ordered phases. The left diagram (Fig. 10a) shows the phase diagram as a function of a scaled reciprocal temperature (i.e. the interaction parameter χ and the composition $\Phi = -1 + 2\phi_{\text{oil}}$. There are different regions indicated by D for disordered, L for lamellar, and H for hexagonal, and further coexistence regions. This phase diagram has a prominent disordered region, which would mean that oil and water do not form separated domains. For polymers this is possible as we will see later in the manuscript. For microemulsions the interaction parameter would be rather large such that mainly ordered phases exist, at least in this sense that oil and water domains are formed. Equation 6 is quite oversimplified to describe the

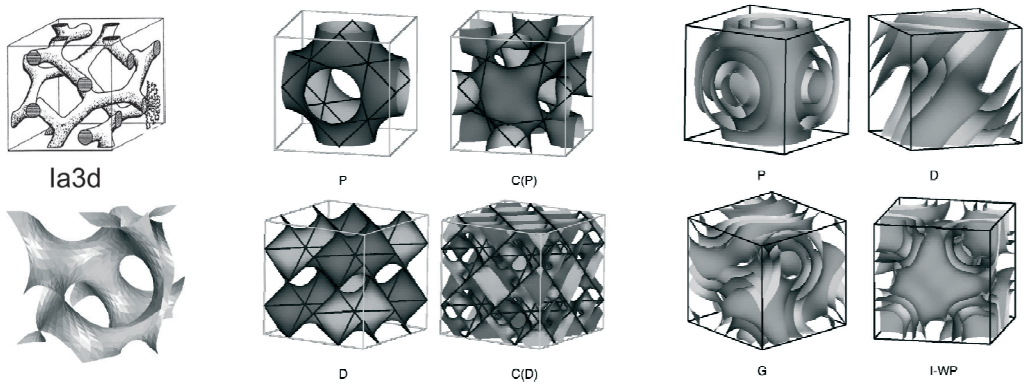


Fig. 9: *Left: The most prominent structure: The gyroid phase G (G_1). Middle: The next most prominent structures in a cubic phase. Especially the D and P (V_1) are realized. Right: Double frequency structures. These structures are not realized often. Graphs from [19, 20].*

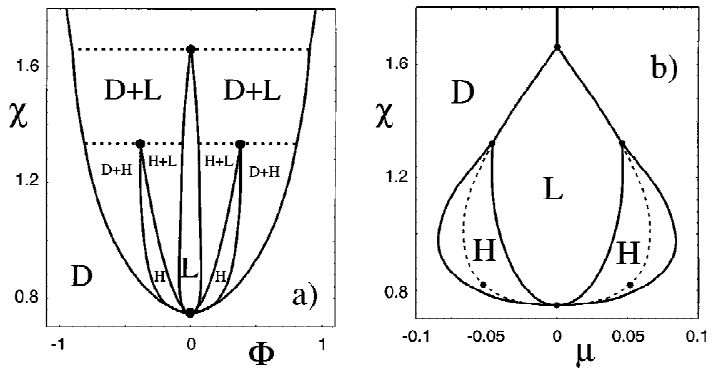


Fig. 10: *Two-dimensional bulk phase diagram [22], showing disordered (D), lamellar (L), and hexagonal (H) phases, as a function of the interaction parameter χ. The x-axis is spanned by (a) the average order parameter $\Phi = -1 + 2\phi_{\text{oil}}$, and (b) the chemical potential μ. Dashed lines in (a) denote triple lines and dashed lines in (b) denote the (metastable) L-D transitions, which exhibit tricritical points (denoted by solid circles).*

complex behavior of microemulsions. So there exist more detailed approximations (see Ref. [24, 25]), which aim at better descriptions, but on the other hand the more complicated algebra cannot be discussed in this manuscript. It should be mentioned that Ref. [25] treats Coulomb interactions quite explicitly.
In summary, similar to the aqueous surfactant systems, we have seen that the Helfrich free energy explains already simple micellar structures with a high degree of symmetry although this concept deviates from the concept of the packing parameter for aqueous surfactant systems. Interactions between neighboring membranes play an important role for liquid crystalline phases. They have translational symmetry in addition. Using a simple order parameter as a long wavelength approach, the Landau expansion is capable of predicting ordered phases in parallel. At this point, the experimental and theoretical access to the underlying coefficients of the different approaches still needs to be made clear.

4 Phase Diagrams of Microemulsions

We have already seen the Gibbs representation of a phase diagram for microemulsions at constant temperature. The next important parameter to be taken into account is the temperature as a third axis. An example for such a phase diagram in displayed in Fig. 11 which is rather realistic for non-ionic surfactants [26]. There are coexistence regions that arise from unfavorable interactions between the components. At high temperatures, the hydrogen bonds are weaker and especially break between the non-ionic surfactant head (ethylene-oxide groups) and water. Thus, the 2-phase region comes in from the top on the water-surfactant face. Similarly, the oil-surfactant tail interactions become unfavorable at low temperatures. In the ultimate limit, they describe the formation of wax crystals due to enthalpic interactions. The third and most important immiscibility persists between oil and water, and stays nearly unchanged for all temperatures. The latter immiscibility gives rise to a big region on the water-oil face. Interestingly, at intermediate temperatures, the surfactant is capable of partially mixing large fractions of water and oil in the three-phase-coexistence region (dark grey) within the surrounding two-phase coexistence regions. Apart from that, towards higher surfactant concentrations, the one-phase microemulsion is observed. Further phases to this end were already discussed before (previous section, see Figs. 4 and 6) and are neglected here.
For many experimental studies and for simplicity reasons, one limits the description of the phase diagrams to two-dimensional plots. If temperature needs to be included, there is the ω-cut that is mostly applied for observing oil in water microemuslions. The fraction $\omega = m_{\mathrm{surf}}/m_{\mathrm{water}}$ between surfactant and water is kept constant, while sequentially more and more oil is added, such that $w_{\mathrm{B}} = m_{\mathrm{oil}}/(m_{\mathrm{water}} + m_{\mathrm{oil}} + m_{\mathrm{surf}})$ is increased. A wider temperature range opens up for the one-phase droplet microemulsion at low oil concentrations, and then becomes narrow towards the high oil concentrations. Using small angle neutron scattering, it can be shown that the oil droplet dimensions increase with increasing w_{B}.
Another cut is called isopleth. It keeps the fraction $\phi = m_{\mathrm{water}}/(m_{\mathrm{water}} + m_{\mathrm{oil}})$ between water and oil constant. The amount of surfactant $\gamma = m_{\mathrm{surf}}/(m_{\mathrm{water}} + m_{\mathrm{oil}} + m_{\mathrm{surf}})$ is sequentially increased such that the solubility between water and oil is increased. First, there is the unimeric solubility of the surfactant (determined by the critical micelle concentration γ_0), then the three-phase coexistence region, and finally the one-phase microemulsion. The scheme of such a ϕ-cut fish phase diagram is displayed in Fig. 12. The observation of different coexisting phases is schematically displayed by the test tubes with two or three phases. The experimental

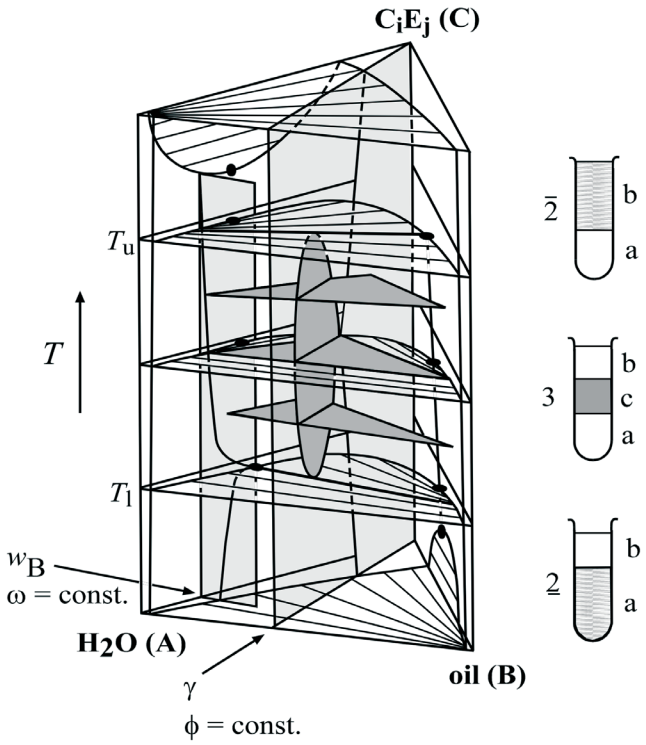


Fig. 11: *The next dimension to the Gibbs phase triangle spans a prism [26]. The bottom still has the classical corners of water, oil, and surfactant. The coexistence regions are displayed in this sketch with many tie lines, and the principal appearances of phases are indicated on the right by* $\bar{2}$ *for an excess water phase with a microemulsion, 3 for three-phase-coexistence of water, oil and a microemulsion, and* $\underline{2}$ *for an excess oil phase with a microemulsion. The cuts through the phase prism (shaded grey) are discussed in the main text.*

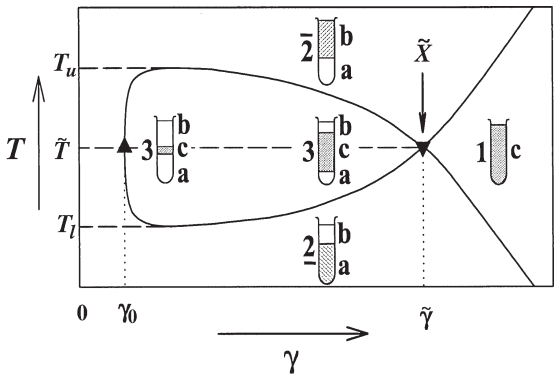


Fig. 12: *The isopleth cut of a microemulsion phase diagram: Temperature versus surfactant content [27]. The two- and three-phase coexistence are indicated by* $\bar{2}$, $\underline{2}$ *and 3 with the appearance of phases in a test tube. The phases are (a) excess water, (b) excess oil, and (c) microemulsion. Finally, at highest surfactant amounts, the one-phase microemulsion (1) is found.*

determination of the three-phase coexistence boundary (fish body) is exactly observed by the coexistence of the three phases, which disappear differently in different directions of the phase diagram. The highest and lowest temperatures of the fish body are called T_{u} and T_{l}. Along the vertical axis there is a high degree of symmetry due to the change of the spontaneous curvature c_0 with temperature, which is found best for non-ionic surfactants of the C_iE_j-type. At the phase inversion temperature $\tilde{T}$ the spontaneous curvature c_0 is zero. Here, the solubility between water and oil is found to be best, which can be seen by the one-phase region. It extends to the lowest surfactant amount $\tilde{\gamma}$ at the fish tail point $\tilde{X} = (\tilde{\gamma}, \tilde{T})$. Especially, the one-phase microemulsion with its phase boundaries is of high interest for characterizing the microemulsion system, because (a) the surfactant efficiency is determined by $\tilde{\gamma}$, and (b) the experimental preparation of these systems is easy here due to the thermodynamic stability. Contrarily, the observation of the three-phase coexistence in equilibrium takes days.

The amount of surfactant introduces a certain amount of interface between the water and oil domains. From the elementary understanding of Helfrich's model (Eq. 2) there is an invariance of the phase stability with surfactant amount[2] [28]. This means that at any given surfactant amount the same phase would appear with only the difference of length scales. Experimentally, this is not the case (see Fig. 12). At constant temperature along different surfactant amounts γ there are clearly different phases observed. The reason for this is the thermal fluctuations. A fluctuating membrane does not keep the original orientation of an arbitrary central point. Within the persistence lenght ξ_{p} the membrane is nearly flat, which is also called a patch. This patch size is given by[3]:

$$\xi_{\mathrm{p}} = a \exp(2\pi\kappa/k_B T) \tag{7}$$

with the molecular length a of ca. 1nm. Beyond this distance, the continuous membrane has 'forgotten' about the original orientation and takes a different direction. So the bicontinuous microemulsions consists of many patches that are connected to form a continuous membrane. This concept of persistence lengths is known from polymers, which form random coils on large length scales, but appear locally rigid or rod-like. The rigid step is elementary and is repeated sequentially with differing, random orientations. So the fluctuating membrane is something like a random walk extending to two dimensions.

The view of patches can also be reverted to a renormalized contribution to the bending rigidity that is length scale dependent (i.e. on L). A schematic solving of Eq. 7 for κ would approximately express an additive term like:

$$\Delta\kappa = -\alpha \frac{k_B T}{4\pi} \ln(L/a) = \alpha \frac{k_B T}{4\pi} \ln(\Psi) \tag{8}$$

with $\alpha = 3$ and the membrane volume fraction $\Psi = aS/V = a/L$.[4] Experimentally, the membrane volume fraction $\Psi \equiv \gamma$ is the surfactant content. This finding gives a new connection to the γ axis in Fig. 12, the isopleth cut. On the one hand, when changing the surfactant content within the one-phase region the effective bending rigidity would change. We will see below that there is experimental evidence for such an interpretation. For the fish body, there are tie lines for the microemulsion phase towards the right boundary, or more relaxed to the fish tail point.

[2] The vertical axis of Fig. 8 is $\omega = V/(SR_0)$ which also contains the scale invariance of the Helfrich energy with scaling the size R_0.

[3] This relation was derived for a slightly excited flat membrane using the Helfrich energy and the equipartition theorem between different modes. The latter looks on undulation modes in reciprocal space.

[4] The surface per volume S/V can in principle be observed by scattering experiments.

Thus, all microemulsion phases in the three phase coexistence have (nearly) the same surfactant concentration $\sim \tilde{\gamma}$. I.e. only when the minimum of surfactant concentration is reached, the system can transit to the one-phase microemulsion and change the domain sizes accordingly. Or the surfactant is only capable of stabilizing a maximum domain size. Finally, there are more phases observed at higher surfactant amounts (see Fig. 6), which also arise from the renormalization dependent bending rigidities.

To summarize this section, we have seen two most important cuts for phase diagrams of microemulsions. The ω-cut was useful for droplet microemulsions, and the fish phase diagram was useful for bicontinuous micreoemulsions. In the case of non-ionic surfactants of C_iE_j-type, the coordinates, especially for the fish phase diagram, are easily described. Here, we saw that the temperature axis is simply connected to the mean curvature c_0. The surfactant amount gives rise to a renormalization of the bending rigidity and takes the scale invariance away. Only then, the different phases along the surfactant amount axis can be explained theoretically. This view on thermal fluctuations is a little different from the effective interactions between neighboring, fluctuating membranes through the Helfrich interaction. In either way, fluctuations are highly important to explain the phase behavior of microemulsions.

5 Polymer Boosting Effect

The polymer boosting effect was first observed by phase diagram measurements [29, 30]. For characterizing the efficiency of a surfactant equal amounts of water and oil are mixed with a variable amount of surfactant $\phi_{\mathrm{C}} \equiv \gamma$. Then the phases are determined as a function of temperature for each $\phi_{\mathrm{C}} \equiv \gamma$. Such a phase diagram is displayed in Fig. 13. Without polymer there is a one-phase region (fish tail) with a minimum amount of surfactant, which is needed to solubilize the water and oil. This surfactant amount is a characteristic figure for the efficiency of a surfactant. When adding polymer as a fourth component the total amphiphile concentration $\phi_{\mathrm{C+D}}$ is the considered variable. The relative amount of the polymer is given in units $\delta = m_{\mathrm{polymer}}/(m_{\mathrm{surf}} + m_{\mathrm{polymer}})$, which takes values of 1.4 to 10%. However, the absolute values in the overall microemulsion are tiny and take values from 0.2 to 0.4%. Nonetheless, these small amounts of polymer are responsible for the one-phase region to move to smaller amphiphile concentrations. This means the diblock copolymer makes the surfactant more efficient. Using small angle neutron scattering experiments under contrast variation [31] it could be proved that the diblock copolymer is anchored in the surfactant membrane. So each block finds the way in the domain where it is soluble, and takes a mushroom-like conformation. Due to its anchoring the polymer exerts a pressure on the membrane, which is responsible for an effectively higher membrane rigidity. This leads to the formation of larger domains with a better surface to volume ratio. This is the quick explanation for the polymer boosting effect, which we shall discuss in more details now.

We have already seen that the free energy of a microemulsion is dominated by the elastic behavior of the membrane [16]. There are two moduli κ and $\bar{\kappa}$, which describe the energy needed to deform the membrane with a certain mean curvature and a saddle splay curvature. For simplicity, we assume that the equilibrium curvature c_0 is zero, which is true for the phase inversion temperature, the temperature of the fish tail point. The bending rigidity depends on different physical contributions as we will see now [32]:

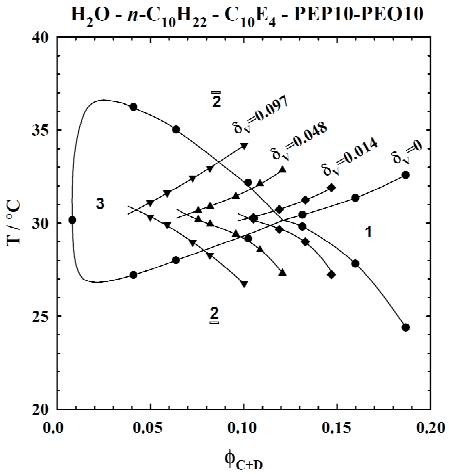


Fig. 13: *Phase diagram: Temperature as a function of the amphiphile content* $\phi_{\mathrm{C+D}}$. *The phase diagram without polymer (•) shows different regions: At high and low temperatures there are two-phase coexistence regions (*$\bar{2}$ *and* $\underline{2}$*). For intermediate temperatures at higher surfactant contents there is the one-phase bicontinuous microemulsion (1). For intermediate temperatures at low surfactant contents there is the three-phase coexistence region (3) with a microemulsion coexisting with a water-rich and an oil-rich phase. Furthermore, there are the one-phase boundaries (fish tails) shown for additions of amphiphilic diblock copolymer at concentrations of* $\delta = 0.014(\blacklozenge), 0.048(\blacktriangle)$ *and* $0.097(\blacktriangledown)$. *The added polymer was* PEP_{10}-PEO_{10}.

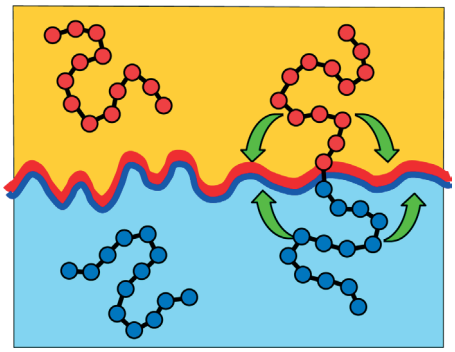


Fig. 14: *Scheme of homopolymers and diblock copolymers at a surfactant membrane in a microemulsion. The homopolymer favors membrane fluctuations while the diblock copolymer exerts a pressure on the membrane, which causes flattening.*

$$\begin{aligned}
\frac{\kappa_R}{k_BT} &= \frac{\kappa_0}{k_BT} && \text{pure membrane}\\
&+ \frac{\alpha}{4\pi}\ln(\psi) && \text{thermal fluctuations}\\
&- \beta\frac{\phi_p(R_{hW}^3 + R_{hO}^3)}{V_p} && \text{homopolymer}\\
&+ \Xi\sigma(R_{dW}^2 + R_{dO}^2) && \text{diblock copolymer}
\end{aligned} \tag{9}$$

The first contribution arises from the membrane itself. The surfactant molecules withstand deformations due to their molecular structure. The next addend describes the spatial renormalization. Due to the fluctuations of the membrane the membrane looks less rigid on larger length scales ($\alpha = 3$). The negative sign arises from the logarithm of the membrane volume fraction $\psi = \phi_{\mathrm{C}} - 0.01 < 1$, which is the total surfactant content minus the unimerically dissolved sur-

factant. While corrugated paper looks more stiff on larger length scales the membrane shows the opposite effect. The next contribution describes the homopolymer effect[5]. It is proportional to the homopolymer concentration ϕ_p, the cubed end-to-end distances of the water and oil soluble polymers R_{hW} and R_{hO} and the reciprocal volume of the polymer V_p. The last addend is the diblock copolymer contribution[6]. It is proportional to the grafting density σ (no. of polymers per membrane area) and the squared end-to-end distances of the water and oil soluble blocks R_{dW} and R_{dO}. The theoretical effect is also depicted in Fig. 14 where diblock copolymers exert a pressure on the membrane, which leads to flattening. Contrarily, homopolymers facilitate the fluctuations of the membrane. The saddle splay modulus in principle has the same dependency as κ, according to:

$$\frac{\bar{\kappa}_R}{k_B T} = \frac{\bar{\kappa}_0}{k_B T} + \frac{\bar{\alpha}}{4\pi}\ln(\psi) + \bar{\beta}\frac{\phi_p(R_{hW}^3 + R_{hO}^3)}{V_p} - \bar{\Xi}\sigma(R_{dW}^2 + R_{dO}^2) \tag{10}$$

It should be emphasized that the three contributions from fluctuations ($\bar{\alpha} = -10/3$), and the polymers are very similar in magnitude but they have opposite signs as for κ.[7] So one can roughly say that κ and $\bar{\kappa}$ have the same value but opposite signs. The moduli now have to be connected to observable effects in order to compare them. From molecular dynamics simulations the saddle splay modulus takes a certain value at the fish tail point, i.e. $\bar{\kappa}_R = \bar{\kappa}_{\mathrm{FTP}}$. This value is much smaller than the intrinsic surfactant molecule contribution $\bar{\kappa}_0$, and so $\bar{\kappa}_{\mathrm{FTP}}$ can be neglected. So equation 10 can be solved for the surfactant content, which will read then:

$$\psi = \psi_0 \exp\left(\hat{\beta}\frac{\phi_p}{V_p}(R_{hW}^3 + R_{hO}^3) - \hat{\Xi}\sigma(R_{dW}^2 + R_{dO}^2)\right) \tag{11}$$

The minimum surfactant concentration of the pure system arises from the constant $\bar{\kappa}_0$, and can be measured directly. The coefficients $\hat{\beta}$ and $\hat{\Xi}$ are derived from equation 10 by dividing by $\bar{\alpha}$. It is directly obvious that adding diblock copolymer leads to smaller amounts of surfactant needed to solubilize oil and water while homopolymers show the opposite effect. Thus, starting from the Helfrich free energy we have explained how the polymer boosting effect works. But we are still left with the connection of κ to experiments. If one conducts small angle neutron scattering experiment on bicontinuous microemulsions one observes typical scattering curves as depicted in Fig. 15. There is a pronounced peak at a scattering vector q^*, which is connected with the domain spacing $d \approx 2\pi/q^*$. The width of the peak is proportional to the reciprocal correlation length ξ. At small angles there is still considerable forward scattering. So, the microemulsion does not only have alternating domains with a periodicity d, but also long range fluctuations. This arises from local enrichments of water or oil because the surfactant does not fully make sure that the local concentration is the overall concentration. The forward scattering is also directly proportional to the reciprocal osmotic compressibility. At large q there is the Porod law $I(q) \propto Pq^{-4}$, which comes from the sharp surfaces of the water and oil domains. The Porod constant P is proportional to the surface per volume $S_{\mathrm{tot}}/V_{\mathrm{tot}}$, and, therefore, is proportional to the membrane volume content ψ. The overall scattering function is well described by the following formula:

[5] In theory $\beta = 0.0238$.
[6] In theory $\Xi = (1 + \pi/2)/12$.
[7] In theory $\bar{\beta} = 0.0211$ and $\bar{\Xi} = 1/6$.

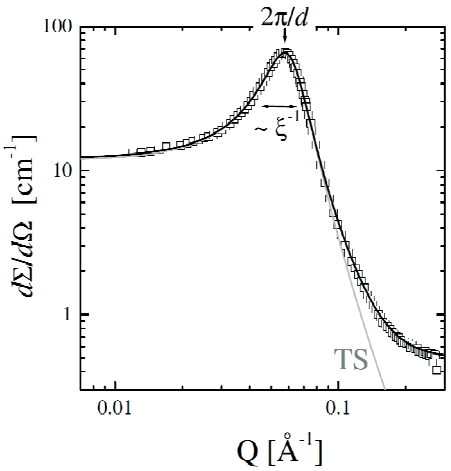


Fig. 15: *A typical scattering pattern of a bicontinuous microemulsion (intensity vs. q) in bulk contrast, i.e. with D_2O, hydrogenated oil and hydrogenated surfactant [34]. From the peak position the domain spacing is derived, while the peak width indicates the correlation length. The grey curve shows the fit with the Teubner-Strey theory. The solid line is a fit with the extended theory of equation 12.*

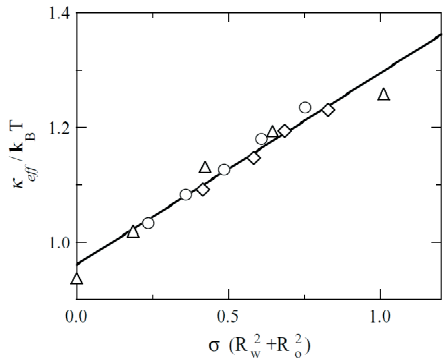


Fig. 16: *The bending rigidity κ as a function of the scaled diblock copolymer amount [30]. According to equation 9 this function is linear. It shows that the membrane becomes more rigid with the diblock copolymer addition. Different symbols arise from different molar masses of the polymer.*

$$I(q) = \frac{d\Sigma}{d\Omega}(q) = \left(\frac{8\pi\langle\nu^2\rangle/\xi}{q^4 - 2(k_0^2 - \xi^{-2})q^2 + (k_0^2 + \xi^{-2})^2} + \frac{G \cdot \mathrm{erf}^{12}(1.06qR_g/\sqrt{6})}{1.5q^4R_g^4} \right) \exp(-\sigma^2 q^2) + b_{\mathrm{backgr}} \quad (12)$$

The normalized intensity $I(q)$ is given by the macroscopic scattering cross section $d\Sigma/d\Omega$. The first fraction in the top line of equation 12 describes the long wavelength behavior for wavelengths down two approximately q^*, i.e. the domain spacing. This expression is known as the Teubner-Strey theory [33]. The term arises from a Landau description of the order parameter, similar to eq. 6. The Landau approach assumes that the free energy can be described as a functional expansion of the order parameter(s). From symmetry considerations, and considerations about the highest order terms needed, one usually arrives at rather simple expressions. Using the Fluctuation-Dissipation Theorem the scattering function can be calculated from the free energy. This basically leads to the fourth order polynomial in the denominator. From the real space correlation function it then can be judged, which structural information is found in the coefficients [31]. Here the real wave number $k_0 = 2\pi/d$ appears, which is only approximately the peak position q^*. The correlation length ξ is also well defined now. The numerator is connected to the scattering length density difference $\Delta\rho$ and the water-water correlation average, according to $\langle\nu^2\rangle = (\Delta\rho)^2\phi_W(1-\phi_W)$ with ϕ_W being the water content. The second



Fig. 17: *Clou Universal paint remover developed using neutron small angle scattering is available in do-it-yourself shops.* © *Clou GmbH*

fraction in equation 12 describes additional surface [34], which is not expressed by the Landau approach, which is obvious because the approach comes from long wavelengths and does not cover the exact domain structure. So the sharp transition from water $(+1)$ to oil (-1) is not well described, and the short wavelength fluctuations shorter than the domain spacing are not well covered either. The expression is rather phenomenological, but was motivated in another context with fractal structures by Beaucage. His approach described the long wavelength behavior by a Guinier approach, and the short wavelength behavior was exactly this term we find here, except that we restricted ourselves to the Porod behavior for sharp surfaces. Here, the radius of gyration R_g describes the size of a single domain (i.e. $R_g \sim d/2$). The amplitude G is correlated with the amount of additional surface while the overall Porod constant is given by $P = 8\pi\langle\nu^2\rangle/\xi + G/(1.5R_g^4)$. The error function $\mathrm{erf}(x)$ is connected to the integral of a Gauss peak. In case that the surfactant molecules are slightly excited individually, the exponential factor takes care of this. Usually, this kind of roughness is described by a length of $\sigma = 2$Å, which is practically invisible for most of the examples. The last addend describes the incoherent background. Mostly, the scattering curves are measured for large enough q, such that the constant level b_{backgr} is well defined. From the scattering experiment we obtain the structural parameters k_0 and ξ. The Gaussian random fields theory [31] connects the structural parameters with the bending rigidity according to:

$$\frac{\kappa_R}{k_B T} = \frac{5\sqrt{3}}{64} k_0 \xi \tag{13}$$

Within the derivation the assumption was made that the bending rigidity is large enough, otherwise a more complicated function will appear. From practical applications formula 13 appeared quite precise [31]. Now, the obtained bending rigidity κ can be compared with the model (see Fig. 16). We obtain a linear increase as a function of the scaled polymer amount. This means that diblock copolymers stiffen the membrane. From literature [31] it is known that the logarithm of the minimum surfactant amount shows the same linear behavior according to eq. 11. This shows that two different observations (scattering and phase diagram) can be compared on the same level through the microscopic interpretation via the Helfrich free energy.

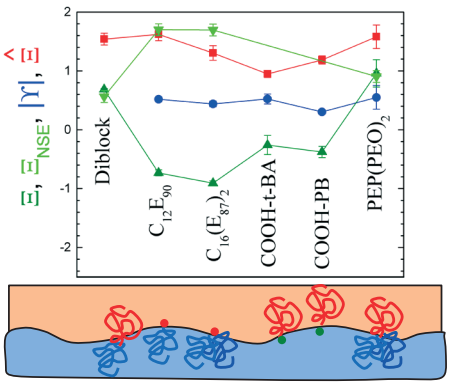


Fig. 18: *The polymer sensitivities* Ξ, Ξ_{NSE}, Υ, *and* $\hat{\Xi}$ *on the parameters* κ, κ_{NSE}, c_0, *and* $\bar{\kappa}$ *of the Helfrich free energy for different symmetric and asymmetric polymers (indicated below).*

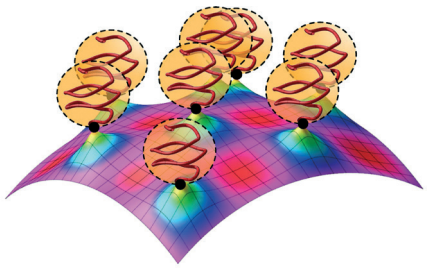


Fig. 19: *The interpretation for the discrepancy of the coefficients* Ξ *and* Ξ_{NSE} *in the case of asymmetric polymers. There are small scale pinches that cause a negative* Ξ, *and accretion of polymers to the bending of the membrane that elevates* Ξ_{NSE}.

While diblock copolymers are quite expensive for industrial applications a simpler way for synthesizing amphiphilic polymers was found. Starting from a linear long alcohol (like dodecanol) the water soluble block can be polymerized quite easily. This yields an amphiphilic polymer with a short hydrophobic part (C_{12}) and a polymer-like water soluble part. This kind of polymer is cheap to produce and it is highly water soluble. The latter is important for formulations because the symmetric diblock copolymer dissolves only very slowly. The slight asymmetry results in a small equilibrium curvature, which is compensated by slightly higher temperatures. So, even for applications a suitable polymer was found and a real product was brought to the marked: The paint remover *Clou* (see Fig. 17).

Apart from the application, several asymmetric polymers were analyzed in more detail using SANS, neutron spin echo spectroscopy (NSE), and phase diagram measurements [35]. The technique NSE measures the relaxations of thermal excitations of the membranes and, therefore, gives another access to the bending rigidity κ. The experimental results for the polymer sensitivities Ξ, Ξ_{NSE}, Υ, and $\hat{\Xi}$ on the parameters κ, κ_{NSE}, c_0, and $\bar{\kappa}$ are displayed in Fig. 18. One observes rather constant coefficients Υ and $\hat{\Xi}$. For the coefficients Ξ and Ξ_{NSE} there is agreement for the symmetric diblock copolymer and the rather symmetric Y-shaped polymer with two hydrophilic arms and one hydrophobic arm. The asymmetric polymers cause a splay of the two coefficients Ξ and Ξ_{NSE}. The explanation is the renormalization of the observed bending rigidity using SANS and NSE. For NSE the membrane patches are big enough to host several polymers, and the polymers are enriched at the preferred curvature. While for SANS the polymers cause additionally small pinches in the membrane and pretend stronger fluctuations that finally lead to a decreased κ within the SANS experiment. The first renormalization is thermodynamically real, and agrees with the unchanged polymer boosting of asymmetric polymers. The second renormalization is an experimental effect that possibly does not play a major role.

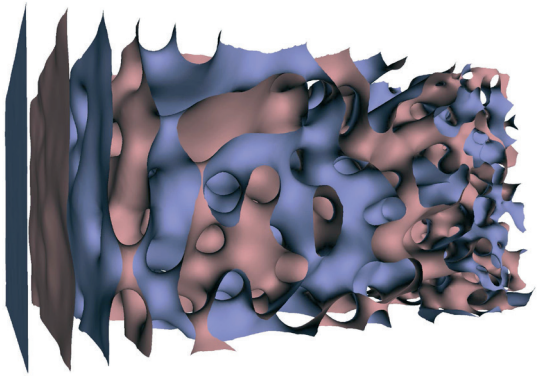


Fig. 20: *A real space image of a microemulsion near a planar hydrophilic wall from a computer simulation [36]. The surfactant layer is depicted with blue and red facing the water and oil domains. Close to the surface a lamellar order is formed while in the volume the microemulsion is bicontinuous.*

6 Microemulsions Near Planar Walls

Surfaces are highly important for the application of microemulsions. This is obvious for cleaning processes because the fluid shall take up the dirt from the surface. But also in enhanced oil recovery applications there are huge surfaces from the sand stone where the oil is located. For instance the cracking fluid is an aqueous surfactant system with wormlike micelles. The micelle network leads to a high viscosity. With this high viscosity the pressure energy can be deposited in the sand stone, which leads to crack formation. To the cracks sand particles (the proppant) are transported to avoid the collapse of the cracks after the application. The aqueous surfactant solution forms a microemulsion in contact with oil, which has a low viscosity. After the application oil can be produced at a higher speed.

So, one important model system to study is a bicontinuous microemulsion adjacent to a hydrophilic planar wall [36]. This question was addressed by computer simulations [36]. A real space picture is shown in Fig. 20. One can see the lamellar order near the surface and the bicontinuous microemulsion in the volume. A kind of order parameter is obtained by laterally averaging the structure as a function of the depth (Fig. 21). Here, two perfect lamellae are observed before the order decays into the volume where an average is reached. This decay is what one would expect for a lamellar order induced by a surface. The real question is how the decaying order of the lamellae is realized. From the lateral cuts in the bottom of Fig. 21 we see that there are perforations in the lamellae, which lead to the decreasing order.

A microemulsion was studied by grazing incidence small angle neutron scattering (GISANS) and reflectometry experimentally. The reflectometry measurements basically confirm the decaying order parameter of the simulations. The GISANS experiments are also sensitive to the lateral structures and so there were contributions from the bicontinuous region as well (Fig. 22). Small angle scattering with grazing incidence leads to an evanescent (tunneling) wave in the sample. So the sample is illuminated with a variable depth. This depth depends on the scattering length density difference of the silicon block, which provides the solid-liquid surface and the overall microemulsion. Furthermore, the incident angle allows for fine-tuning the penetration depth of the evanescent wave. In the current study the penetration depth Λ was

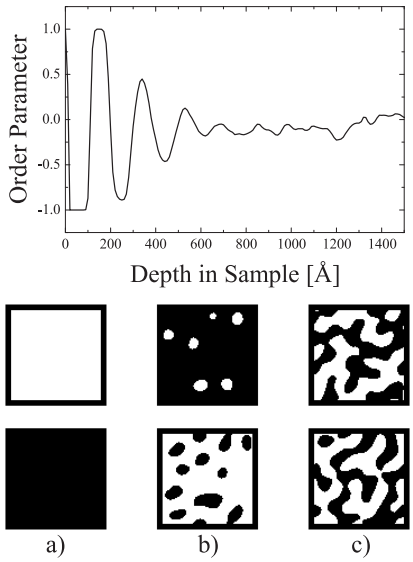


Fig. 21: *Top: The laterally averaged structure of the microemulsion near a planar wall. This function looks like an order parameter of a decaying lamellar order. Bottom: Lateral cuts in different depths (< 100, ~ 300, and 1000Å). There is a) perfect lamellar order, b) perforated lamellae, and c) bicontinuous microemulsion.*

varied between ca. 400 and 1000Å. For small Λ the surface scattering dominates the signal, and the lamellar structure appears only weakly with a Bragg peak. At intermediate $\Lambda \approx 660$Å the Bragg peak becomes more prominent. At higher Λ the bicontinuous microemulsion becomes visible as well. From this experiment the integral intensities of the Bragg peak and the Debye-Scherrer ring are determined. Their ratio is plotted in Fig. 23. The experimental points show an increasing linear behavior from penetration depths of 400Å on where the bicontinuous phase starts to be visible. So the well ordered lamellar phase covers the first 400Å. For the computer simulations the same plot shows that the characteristic depth is ca. 200Å. From the real space structure it is known that from this depth on the perforated lamellae expand. The reason is that the typical length scale of the perforations is nearly the same as for the bicontinuous structure (see Fig. 21). So the GISANS experiment determines the beginning of the perforated lamellae because it appears like an isotropic structure.

The whole concept of depth resolved scattering experiments was transferred to NSE, a spectroscopic method, in oder to observe the thermal excitations of the microemulsion near a planar wall [37, 38]. The highlighting of a certain depth using the evanescent (tunneling) wave is the same as for GISANS. The NSE method shows the relaxation of the structure in the time domain, and delivers a typical time, the relaxation time τ - in principle as a function of the scattering vector Q. Due to very low intensities in this experiments, we limited ourselves to a single $Q = 0.08\text{Å}^{-1}$. The results of the microemulsion, that we considered in this section so far, is displayed by orange symbols in Fig. 24. We see that the relaxation time is slower in the bulk than close to the surface (approx. 3 times accelerated). All intermediate relaxation times can be interpolated using the intensity ratio of Fig. 23. This means, that in the dynamic experiment the same structure is observed that in the static GISANS experiment.

The acceleration is understood by the confinement of the enclosed water volume between the

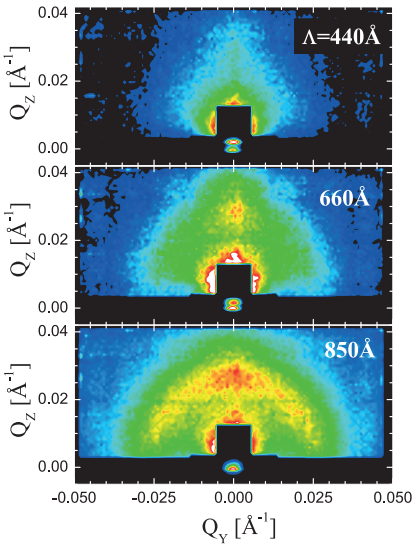


Fig. 22: *GISANS patterns at different penetration depths Λ. For 440Å there is a rather strong surface scattering background in the center and a lamellar peak is only slightly indicated in the middle top. For 660Å the lamellar peak becomes stronger. For 850Å both the lamellar peak and the bicontinuous Debye-Scherrer ring are visible.*

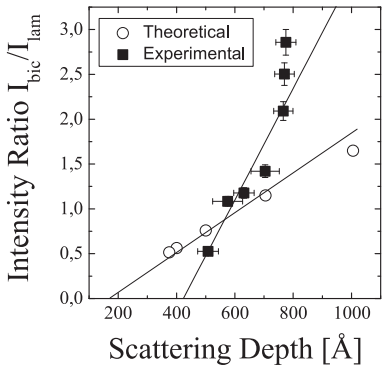


Fig. 23: *The integral intensity ratio of the bicontinuous and the lamellar structure as obtained from GISANS experiments (■). At a penetration depth (scattering depth) Λ of ca. 400Å the ratio starts to grow linearly. This value indicates the beginning of the perforated lamellae. For the simulations (○) the ratio starts to grow already at 200Å where the perforated lamellae are found explicitely.*

wall and the first membrane. This confinement leads to a different, faster hydrodynamic response to the membrane close to the wall. This accelerated feedback is directly observed as a faster relaxation time. The principal concept was elaborated by Seifert, and is discussed in more detail in Ref. [37].

Another question arouse for the polymer additive: How does a boosted microemulsion behave at a planar wall? For this, we used the symmetric diblock copolymer and studied the structure and dynamics employing grazing incidence SANS and NSE experiments [38]. For the static structure we observed that there is a polymer enrichment close to the wall. The dynamics showed a rather weakly slower relaxation compared to the bulk (Fig. 24). But the splay at small distances from the wall between the polymer free and polymer doped microemulsion is clearly visible. The polymer decorated membrane at the surface is more stabilized with respect to thermal fluctuations, i.e. the membrane is simply more rigid. For a more rigid membrane it is simply more difficult to move in a hydrodynamic environment, and therefore the relaxation time is reduced.

The practical benefit of this study is the polymer enrichment at the solid surface. The emul-

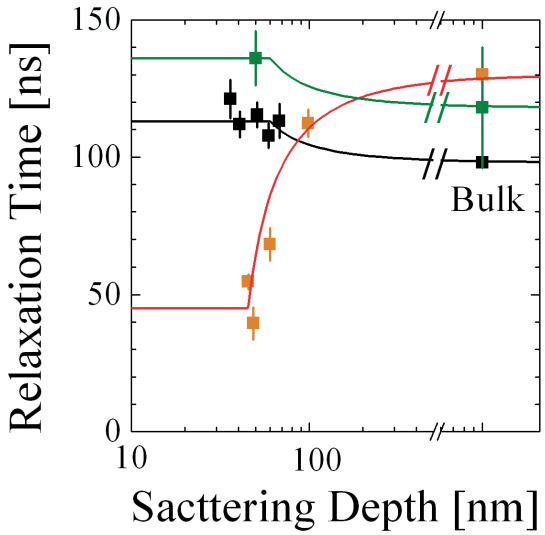


Fig. 24: *The relaxation times of a microemulsion adjacent to a planar wall as a function of the scattering depth. The orange symbols correspond to the polymer free microemulsion. The interpolation line arises from the scattering intensities of Fig. 23. The green and black symbols arise from a microemulsion with a symmetric diblock copolymer additive carried out on different NSE instruments.*

sification ability is increased due to the boosting effect and means that contaminants can be removed from solid surfaces more easily. In applications like soil cleaning (environmental soil remediation) this effect could be highly beneficial.

7 Summary

We have described the theoretical concepts of microemulsions that base on the ideas of Helfrich. A very suitable additive was identified: an amphiphilic polymer that anchors to the surfactant membrane. Especially, the asymmetric polymer is extremely suitable for applications and a first product is on the marked using this additive. Looking at microemulsions adjacent to solid surfaces reveals additionally many interesting aspects that are important for cleaning processes, for instance the soil remediation.

References

[1] W.F.C. Sager, *Curr. Opin. Colloid Interf. Sci.* **3**, 276 (1998)

[2] M. Schwan, L.G.A. Kramer, T. Sottmann, R. Strey, *Phys. Chem. Chem. Phys.* **12**, 6247 (2010)

[3] I. Kayali, K. Qamhieh, U. Olsson, L. Bemert, R. Strey, *J. Disp. Sci. Techn.* **33**, 516 (2012)

[4] G.J.T. Tiddy, *Physics Reports* **57**, 1 (1980)

[5] J.S. Keiper, R. Simhan, J.M. DeSimone, G.D. Wignall, Y.B. Melnichenko, H. Frielinghaus, *J. Am. Chem. Soc.* **124**, 1834 (2002)

[6] A.-G. Fournial, Y. Zhu, V. Molinier, G. Vermeersch, J.M. Aubry, N. Azarouali, *Langmuir* **23**, 11443 (2007)

[7] J.N. Israelachvili, D.J. Mitchell, B.W. Ninham, *J. Chem. Soc. Faraday Trans. I* **72**, 1525 (1976)

[8] J.S. Pedersen, S.U. Egelhaaf, P. Schurtenberger, *J. Phys. Chem.* **99**, 1299 (1995) and G. Jerke, J.S. Pedersen, S.U. Egelhaaf, P. Schurtenberger, *Phys. Rev. E* **56**, 5772 (1997)

[9] R. Strey, R. Schomäcker, D. Roux, F. Nallet, U. Olsson, *J. Chem. Soc. Faraday Trans.* **86**, 2253 (1990)

[10] E. Jahns, H. Finkelmann, *Colloid & Polymer Sci.* **265**, 304 (1987)

[11] R. Strey, *Phys. Chem. Chem. Phys.* **97**, 742 (1993)

[12] H.T. Davis, J.F. Bodet, L.E. Scriven, W.G. Miller, *Physica A* **157**, 470 (1989)

[13] Internal communication with M. Kraus, G. Goos, and G. Gompper; see also G. Gompper, M. Kraus, *Phys. Rev. E* **47**, 4301 (1993) and G. Gompper, G. Goos, *Phys. Rev. E* **50**, 1325 (1994)

[14] R.G. Larson, *J. Chem. Phys.* **91**, 2479 (1989)

[15] N.R. Washburn, T.P. Lodge, F.S. Bates, *J. Phys. Chem. B* **104**, 6987 (2000)

[16] W. Helfrich, *Z. Naturforsch., A: Phys. Sci.* **33**, 305 (1978)

[17] S.A. Safran, L.A. Turkevich, P. Pincus, *J. Physique, Lettres* **45**, L-69 (1984)

[18] E.M. Blokhuis, W.F.C. Sager, *J. Chem. Phys.* **115**, 1073 (2001)

[19] U.S. Schwarz, G. Gompper, *Phys. Rev. E* **59**, 5528 (1999)

[20] U.S. Schwarz, G. Gompper, *Phys. Rev. Lett.* **85**, 1472 (2000)

[21] T. Schilling, O. Theissen, G. Gompper, *Eur. Phys. J. E* **4**, 103 (2001)

[22] R.R. Netz, D. Andelman, M. Schick, *Phys. Rev. Lett.* **79**, 1058 (1997)

[23] F. Schwabl, *Statistical Mechanics* (Springer, Berlin, Heidelberg, New York, 2002)

[24] W.B. Lee, R. Mezzenga, G.H. Fredrickson, *Phys. Rev. Lett.* **99**, 187801 (2007)

[25] X.J. Li, M. Schick, *Biophys. J.* **78**, 34 (2000)

[26] R. Strey, *Colloid Polym. Sci.* **272**, 1005 (1994)

[27] T. Foster, dissertation "Microemulsions as compartmentalised reaction media: structural characterization of water-in-oil microemulsions", Cuvillier Verlag Göttingen, ISSN 978-3-86727-115-8, 2007

[28] G. Gompper, lecture B9 "Statistical Mechanics of Membranes" in 33rd Spring School *Soft Matter, Complex Materials on Mesoscopic Scales*, Forschungszentrum Jülich GmbH, Jülich, 2002

[29] B. Jakobs, T. Sottmann, R. Strey, J. Allgaier, L. Willner, D. Richter, *Langmuir* **15**, 6707 (1999)

[30] H. Endo, J. Allgaier, G. Gompper, B. Jakobs, M. Monkenbusch, D. Richter, T. Sottmann, R. Strey, *Phys. Rev. Lett.* **85**, 102 (2000)

[31] H. Endo, M. Mihailescu, M. Monkenbusch, J. Allgaier, G. Gompper, D. Richter, B. Jakobs, T. Sottmann, R. Strey, I. Grillo, *J. Chem. Phys.* **115**, 580 (2001)

[32] D. Byelov, H. Frielinghaus, O. Holderer, J. Allgaier, D. Richter, *Langmuir* **20**, 10433 (2004)

[33] M. Teubner, R. Strey, *J. Chem. Phys.* **87**, 3195 (1987)

[34] C. Frank, H. Frielinghaus, J. Allgaier, H. Prast, *Langmuir* **23**, 6526 (2007)

[35] M. Brodeck, S. Maccarrone, D. Saha, L. Willner, J. Allgaier, G. Mangiapia, H. Frielinghaus, O. Holderer, A. Faraone, D. Richter, *Colloid Polym. Sci.*, DOI 10.1007/s00396-014-3449-8 (2015)

[36] M. Kerscher, P. Busch, S. Mattauch, H. Frielinghaus, D. Richter, M. Belushkin, G. Gompper, *Phys. Rev. E* **83**, 030401(R) (2011)

[37] H. Frielinghaus, M. Kerscher, O. Holderer, M. Monkenbusch, D. Richter *Phys. Rev. E* **85**, 041408 (2012)

[38] F. Lipfert, H. Frielinghaus, O. Holderer, S. Mattauch, M. Monkenbusch, N. Arend, D. Richter,*Phys. Rev. E* **89**, 042303 (2014)

BIOMATTER



E 1 Active and Growing Materials

Jens Elgeti
Institute of Complex Systems / Theoretical Soft Matter and Biophysics
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Homeostatic pressure **4**

3 Particle based Modeling **5**

4 Competition **7**

5 Fluid or Solid **8**

6 Spheroids **10**

7 Conclusions **13**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

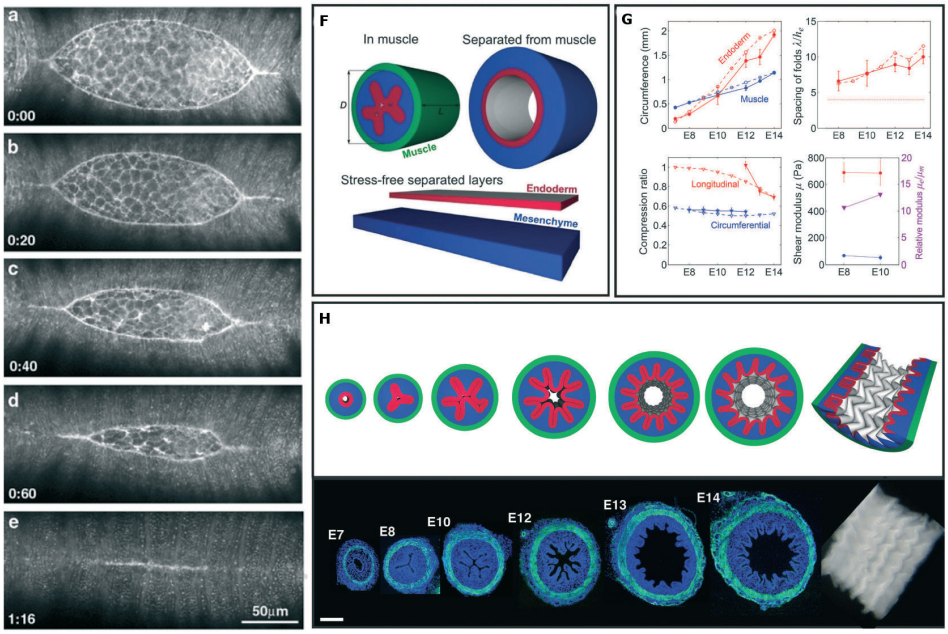


Fig. 1: *Examples of tissue mechanics* – **(a-e)** *Dorsal Closure in the development of Drosophila. The inner tissue (amnioserosa) contracts and the outer ephithelium closes the opening. The mechanical force for closure comes from tension in the amnioserosa (increased by apoptosis) and from a super-cellular pursestring contracting at the leading edge (from [1]).* **(F-H)** *Growth of the tissue making up the entestine against an elastic background causes an elastic buckling instability resulting in the typical wrinkled patterns (from [2])*

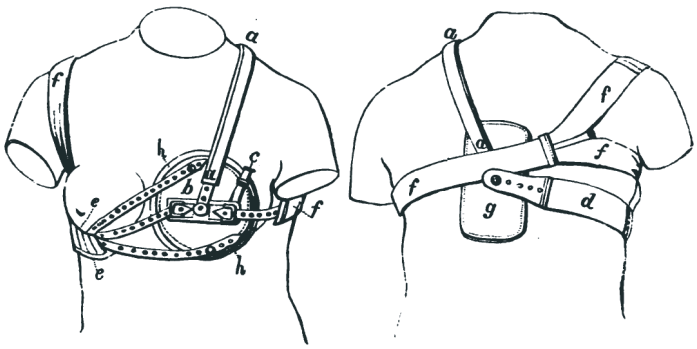
1 Introduction

Materials have classically been studied either in equilibrium, or close to equilibrium. Often, the driving force comes from outside, either in form of moving boundaries as in a shear plate rheometer, or from an external field like gravity.

Active materials are intrinsically out of equilibrium. They consume energy internally on the scale of the microscopic constituents. Materials can be active in many ways. For one, internal forces can create stresses that create spontaneous flows, as in collections of microswimmers (F7) and active gels (F8).

A different route to active materials that we want to address here, is breaking mass conservation. A material that locally generates and destroys itself. Examples are polymerizing gels or cellular tissues. In this lecture, we will focus on cellular tissues, but many aspects are more generic and can be applied to any material that has a stress dependent self generation.

From a biological point of view, the notion has spread that ”Mechanosensitivity in one form or another appears to be a property shared by all cells of the body and by all phyla from mammals to bacteria.”[4] Examples of mechanical feedback range across all organisms found on earth: In the development of an embryo tissues grow and large scale motions of tissues reorganize the growing embryo[5]. Many of these are proven to have a strong mechanical contribution. Dorsal closure in drosophila development is mediated via tensile stresses in the amnioserosa, gastrula-



The slack air-cushion applied to the breast; front and back views.—(a a.) The spring passing over the shoulder, and fixed in front to the centre point of the shield (b), and in the back to the back-pad g.—(b.) The shield, receiving on a screw at its centre point the spring (a), where also is attached a double buckle for straps; round the border of the rim appears an edge (h) of the air-cushion.—(c.) The stop-cock for inflating the air-cushion.—(d.) A belt fixed to the centre point of the pad (g) behind, passing round underneath the healthy breast, and fixed in front by leather straps to the buckle attached to the shield, and also to studs at two opposite points of the circumference of the latter.—(e.) A soft moveable pad attached to the inner surface of that part of the belt (d), which lies below the healthy breast.—(f.) A second belt passing under the arm of the unaffected side by a shoulder strap, carried behind, and fixed to the buckle in front of the frame.—(g.) The counter pressure pad.—(h.) Border of the air-cushion, as before explained. [In order to understand the mode of application of the waist-spring, it is only necessary to suppose a belt substituted for the spring (a), and a spring fixed to the pad (g), passed round the waist in the direction of, and lying outside the belt (d).]

Fig. 2: *Soft mechanical pressure as a possible route for treatment of cancer has tried already in the early 19th century. This device from Walshe in 1846 [3] exerts a "constant, equable, and uniform pressure" on the tumor.*

tion is caused by a spontaneous curvature of a certain group of cells[6]. Also the human adult body undergoes a constant cell turnover. For example our intestine is under constant renewal. Again, mechanics are speculated to be the cause of the observed patterns [7, 2]. Finally, cancer, the "emperor of all maladies" [8] is a disease of growth. A notion that mechanics may play a role and could be used for treatment has appeared early on [3] (See Fig. 2), but only recently received much more attention [9, 10, 11].

Thus, understanding the mechanics of growth can not only help us understand the development of an organism, it will also help to understand and treat diseases or even design artificial tissues. Indeed tissue engineering and organ printing are becoming thriving fields of research. We begin to understand, that mechanical properties of the environment are key players when growing an artificial organ.

In this lecture we will adress some simple key concepts of modeling the mechanics of tissue growth. We begin with a simple continuum model in the spirit of hydrodynamics, followed by a particle based model similarly inspired by particle based hydrodynamics. Subsequently we study growth and competition and mechanical response with those models. Finally, we address an experimental setup that can be well compared with the models.

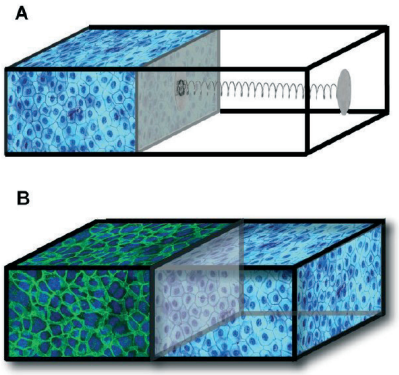


Fig. 3: *Gedankenexperiment to the homeostatic pressure .* **(top)** *A tissue inside a finite compartment permeable to nutrients, but impermeable to cells will grow to fill the available space. If one of the walls is a movable piston connected to a spring, the tissue will press on the spring, until the restoration force is strong enough to hinder further growth. This is the homeostatic pressure.* **(bottom)** *If the spring of the second compartment is replaced by a tissue of different homeostatic pressure, the pressure in the chamber evolves to an intermediate pressure. The weaker cells die, while the stronger tissue grows. Eventually, the stronger tissue completely takes over the compartment.*

2 Homeostatic pressure

When considering the mechanics of growing tissues, growth is the key difference to the otherwise classic mechanics of liquids or solids [12, 13][1]. To account for growth and death, a source term is added to the continuity equation

$$\partial_t \rho + div(\vec{J}) = (k_d - k_a)\rho \tag{1}$$

with k_d and k_a the devision and apoptosis rate respectively and $\vec{J}$ the flux of cells. Here, we will study the effect of mechanical forces on growth. Certainly, growth depends on the biochemical environment, but we will assume this to be constant across the system at hand. Instead, we will ask ourselves which effect mechanical stress has on growth. A simple assumption, is the existence of a homeostatic state, where division and apoptosis are balanced. Homeostasis is the tissue dynamics equivalent of a steady state. We will assume that the homeostatic state is characterized by a homeostatic density or a homeostatic pressure, the two being coupled to each other via the tissues equation of state. Pressure in this context, is the force felt by the cells hindering growth, not to be confused with the hydrostatic pressure. In terms of thermodynamics, it is the conjugate force to cell volume. The homeostatic pressure (or stress) $P_H = -\sigma_H$ and density ρ_H is the pressure (density) at which division is balanced by apoptosis.
We can then expand the growth rates around this homeostatic state [14]

$$\sigma = \chi^{-1}(\rho - \rho_H) + \sigma_H \tag{2}$$
$$(k_d - k_a) = \kappa'(\rho - \rho_H) \tag{3}$$
$$(k_d - k_a) = \kappa(\sigma - \sigma_H). \tag{4}$$

[1] Active stresses by cells, and stresses generated by divisions and apoptosis events can also significantly contribute, and even change the appropriate material laws. We will consider some of those later in the lecture.

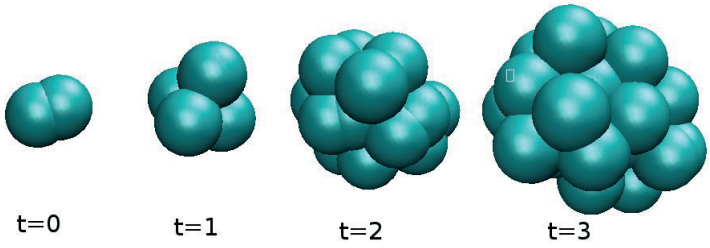


Fig. 4: *Tissues are modeled as collection of sticky soft colloids. Two particles representing a cell expand actively. Once a critical size is reached, the cell divides, creating new particles.*

This implies that a tissue, left in a finite compartment grows to its homeostatic density, exerting its homeostatic pressure on the surrounding (See Fig. 3).

3 Particle based Modeling

Particle based modeling has been used since a long time to model materials in general and fluids in particular. One key concept is to introduce simplified particles and interactions to capture the right dynamics on more mesoscopic length and time scales. For example both multi particle collision dynamics and dissipative particle dynamics (See B3 by R. Winkler) use very simple, but also very different interactions. However, because they both satisfy momentum balance locally, they result in the Navier Stokes equation on larger length-scales. In the same spirit, we introduce a simplified particle based model for tissue growth. One cell is represented by two point particles i and j at positions $\vec{r_i}$, that repel each other with a growth force:

$$\vec{F}^g_{ij} = \frac{B}{(r_{ij} + r_0)^2} \hat{r}_{ij},$$

with a growth strength B. $\vec{r_{ij}} = \vec{r}_j - \vec{r}_j$ is shorthand for the vector connecting particles i and j. This growth force drives the particles constituting one cell apart. Motivated by the "size checkpoint" of real cells, virtual cells divide, once the particles reach a critical distance R_c. For division, two new particles are inserted in the immediate surrounding of the constituents of the mother cell to form the daughters. This is the key active mechanism where mechanics feed back on growth. If the surrounding medium provides additional (mechanical) resistance to growth (due to pressure for example), the process slows down. The second active process is cell apoptosis or death. In the spirit of minimal modeling, apoptosis is introduced by a constant rate of cell removal. In principle this rate could also change through a mechanic feedback, however, this would introduce further parameters.
The rest of the particle based growth model is "classic" soft colloids. A soft repulsive potential of strength f_0 prevents overlap of particles belonging to different cells:

$$\vec{F}^v_{ij} = f_0 \left(\frac{R^5_{pp}}{r^5_{ij}} - 1 \right) \hat{r}_{ij}$$

Where all interactions are set to zero beyond a cut off distance R_{pp}. Adhesion between cells is modeled by a simple constant attractive force

$$\vec{F}^a_{ij} = -f_1 \hat{r}_{ij}.$$

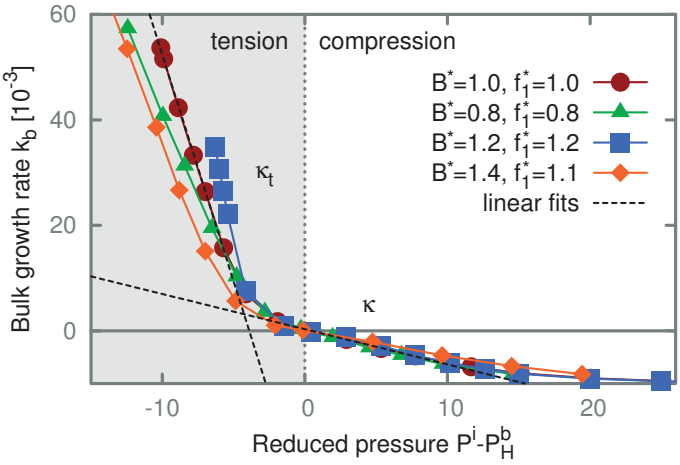


Fig. 5: *Net growth rate $k_b = k_d - k_a$ as a function of the difference between the imposed pressure P^i and the homeostatic pressure P^b_H. Close to the homeostatic pressure, the growth rate is indeed a linear function of the pressure differences. The slope across different parameters is also very similar. However, for larger pressure differences, nonlinearities appear.*

Finally, as in all simulations of active systems, energy has to be dissipated. Here, a Dissipative Particle Dynamics (DPD) -Thermostat (See B3) offers simultaneously momentum conserving dissipation $\vec{F}^d_{ij}$, and random fluctuations $\vec{F}^r_{ij}$.

$$\vec{F}^r_{ij} = \sigma\omega^R(r_{ij})\xi_{ij}\hat{r}_{ij} \tag{5}$$

$$\vec{F}^d_{ij} = -\gamma\omega^D(r_{ij})(\vec{v}_{ij}\cdot\hat{r}_{ij})\hat{r}_{ij}. \tag{6}$$

where σ and γ are the amplitudes of fluctuations and dissipation, $\xi_{ij} = \xi_{ji}$ a symmetric gaussian random variable with zero mean and unit variance and a weight functions $\omega(r_{ij})$ as in DPD. In total, the force between particles reads

$$\vec{F}_i = \underbrace{\vec{F}^g_{ik} + \vec{F}^d_{ik} + \vec{F}^r_{ik}}_{\text{intracellular forces}} + \underbrace{\sum_j \left(\vec{F}^v_{ij} + \vec{F}^a_{ij} + \vec{F}^d_{ij} + \vec{F}^r_{ij}\right)}_{\text{intercellular forces}} + \vec{F}^b \tag{7}$$

On large enough time and length-scales, this model reproduces the concept of homeostatic pressure. A tissue confined by a piston grows to a steady state pressure. When the tissue is compressed, the division rate goes down and apoptosis reduces the number of cells, until the pressure is back at its homeostatic value. Close to the homeostatic pressure, growth rates indeed follow the linear expansion postulated in eq. 2 (See Fig. 5). However, they also indicate strong nonlinearities as the pressure further deviates from the homeostatic one.

Finally, a note on parameters and interpretation. Similar to fluid models, where particles are not representing water molecules, the analogy of two particles describing a cell is just an interpretation to simplify the explanation. When interpreting data from particle based simulations, a simulation cell, could represent many real cells, or vice versa. It is important to match the right mesoscopic quantities when interpreting simulation results in light of experiments.

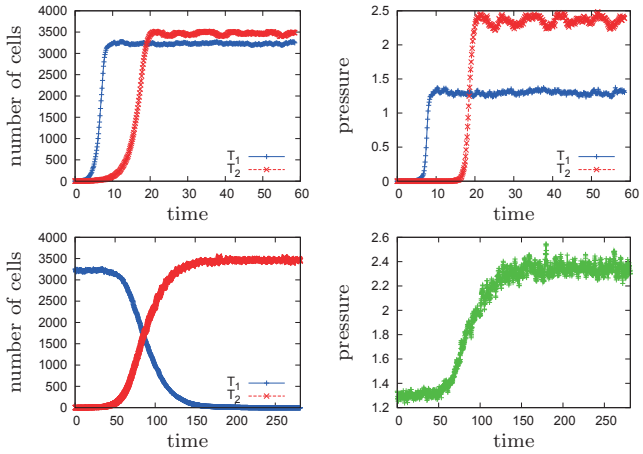


Fig. 6: *Two tissues in a finite compartment face competition for space.* **(top)** *In Isolation, tissue 1 grows faster, but tissue 2 reaches a higher homeostatic pressure.* **(bottom)** *In competition, the compartment reaches a pressure intermediate of the two homeostatic pressures. This causes tissue 1 to shrink, and tissue 2 to grow, until the compartment is completely taken over by tissue 2.*

4 Competition

If two tissues are present in a finite compartment, they enter competition. One of the two tissues will take over the compartment eliminating the other.
Using the theory of homeostatic pressure, we can understand the general dynamics. Assuming the tissue as incompressible and negligible elastic and viscous forces, competition can be expressed by simple number balance. The division rates of both tissues are determined by the stress

$$k_i = \kappa(\sigma - \sigma_{Hi}), \tag{8}$$

where we abbreviated the net growth rate of tissue i as $k_i = (k_{di} - k_{ai})$. Due to incompressibility the compartment has space for a total number $N_{tot} = N_1 + N_2$ cells of tissue 1 and 2. The total number of cells will be constant over time, i.e.

$$0 = \partial_t N_{tot} = k_1 N_1 + k_2 N_2 \tag{9}$$

This is easily solved for the stress

$$\sigma = \frac{N_1 \sigma_{H1} + N2\sigma_{H2}}{N_1 + N_2} \tag{10}$$

and the growth-rates:

$$N_1 k_1 = \kappa \frac{N_1 N_2 (\sigma_{H1} - \sigma_{H2})}{N_1 + N_2} = -N_2 k_2 \tag{11}$$

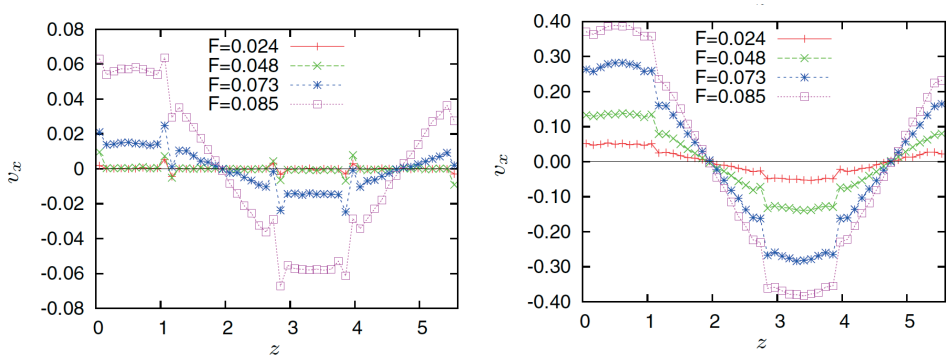


Fig. 7: *Rheology of three dimensional tissues in the particle based model. Results from an in-silico rheology experiment. In a tissue one layer* $0 < z < 1$ *is pulled in the* $x-$*direction, a second layer is pulled in the opposite direction.* **(left)** *Without cell division and apoptosis the tissue behaves like a yield stress solid. Only above a critical stress does the tissue begin to flow.* **(right)** *With cell division and apoptosis, the tissue flows for any given force.*

If we define $\phi = N_1/N_{tot}$ as the fraction of 1-cells and the difference in homeostatic pressure as $\Delta\sigma = (\sigma_{H1} - \sigma_{H2})$, this simplifies to

$$\partial_t \phi = \kappa \Delta\sigma \phi(1-\phi) \tag{12}$$

i.e. a simple logistic growth on time scales $\kappa\Delta\sigma$.
In a particle based simulation, we can study the more complex behavior of a three dimensional competition and more generic parameters. In particular, we can compare tissues that have a higher devision rate if unconstrained, versus tissues with a lower division rate, but a higher homeostatic pressure. As predicted by the analytic model, even in this extreme example, it is the tissue with the higher homeostatic pressure that wins the competition (see Fig. 6). Even though these tissues are neither incompressible nor do they have the same κ, the shape of the curve is still well reproduced by a logistic growth predicted by the simple number balance equation (12).

5 Fluid or Solid

Real tissues have a very complex viscoelastic behavior. On timescales of seconds or minutes, the rheological properties can be measured in a relative straight forward extension of classical rheological experiments. However, growth happens on very long timescales, that are difficult to access experimentally. Extrapolating from short timescales to the long term behavior is difficult, but we can exploit the long timescales to predict some generic features. The timescale of growth, is the longest timescale in the system, much longer than any other cellular timescale like attachment and detachment of cellular adhesion proteins. One can thus already guess that on the timescales of growth, the tissue behaves as a viscous fluid. Even if the tissue behaves like an elastic solid on short timescales, division and apoptosis relax stress, leading to fluid behavior on long timescales. Combining these simple arguments with a simple Maxwell model for viscoelastic behavior results in a rather good prediction for the tissue viscosity η.

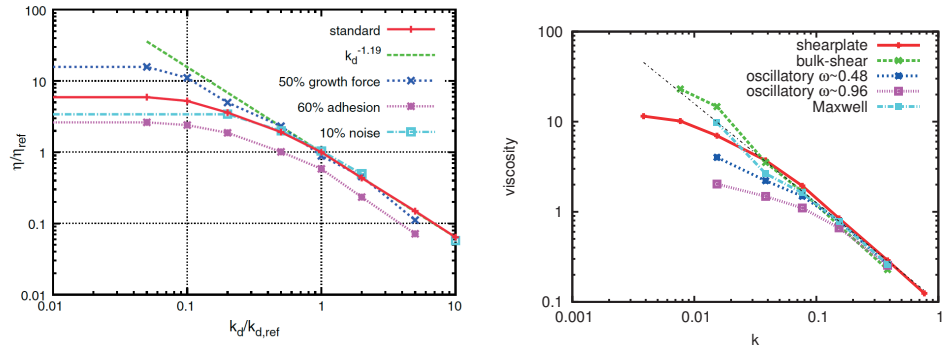


Fig. 8: *Rheology of three dimensional tissues in the particle based model.* **(left)** *Viscosity as determined by a virtual shear plate experiment for different tissue parameters. For cell turnover-rate k larger than the shear rate, the predicted $1/k$ behavior is well reproduced. For smaller cell turnover rates, the yielding behavior leads to a finite plateau of the viscosity* **(right)** *The viscosity can be measured in many different ensembles, which agree with each other when the shear rate is small compared to the cell turnover rate. Fitting a Maxwell model to oscillatory shear flow experiments, yields an long timescale small stress viscosity. This agrees best with the expected $1/k$ behavior.*

In the Maxwell model, the product of the elastic modulus E and the relaxation time τ give the viscosity

$$\eta = E\tau. \tag{13}$$

Arguing that the relaxation time is proportional to the cell turnover time, we get

$$\eta = E\tau \approx E/k \tag{14}$$

with cell turnover $k = k_a = k_d$.
Using the particle based simulations, these arguments can be tested with a very different approach. First, consider a tissue in periodic boundary conditions, one layer is pulled up, and another downwards. A liquid material would display a linear velocity gradient between the two layers, while an elastic material would show no velocity, but a strain gradient. Indeed, if apoptosis is disabled by setting the apoptosis rate $k_a = 0$, division stops automatically when the finite compartment is densely packed with cells. In this situation, the particles behave like a colloidal glass. Below a critical shear stress, no continuos motion can be observed. Above a critical stress, the tissue yields, and starts to flow with rather small velocities (See Fig. 7). With apoptosis, the tissue is in a homeostatic state with continuos cell turnover $k = k_a = k_d$. Each division and apoptosis event locally relax some stress, leading to a viscous behavior.
To really understand the rheological behavior different types of rheology experiments have to be performed. The simplest one is the shear plate experiment: The tissue is confined between two parallel plates, and the top one is moved by a constant velocity. This imposes a shear on the tissue. If the shear rate is larger than the cell turnover rate, the rheology is dominated by the yielding behavior described above for no cell turnover. The effective viscosity is dominated by the yield-stress. If the cell turnover-rate k is larger than the shear rate, the stress-relaxation via cell-turnover becomes important, and the effective viscosity is well predicted by Eq.(14).

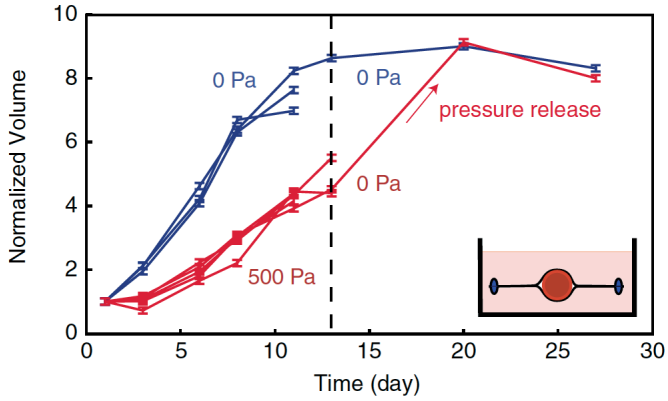


Fig. 9: *Growth of tissue spheroids in dialysis bags. Cells from a colon carcinoma cell line (CT26) are placed inside a dialysis bag. Large molecular weight dextran is added to the growth medium to exert an osmotic pressure via the bag on the spheroids. After 12 days most spheroids are sacrificed for cryosections. Two spheroids are allowed to continue to grow. They grow to the same final volume after pressure is released. From [15]*

On a more quantitative level, oscillatory shear experiments can be performed. A simple way to implement these, is to impose an oscillating (in space and time) force density. From these, the full complex viscosity η_C can be extracted for different frequencies.

6 Spheroids

Experimentally measuring the mechanical properties of tissues on long timescales is very challenging indeed. The tissue has to experience the same mechanic stimulus over the course of many days or weeks, while keeping the biochemical environment constant. One such experiment is growing tissue spheroids under isotropic osmotic stress. Cells are placed in a dialysis bag. The bag is put in a standard growth medium, with an added high molecular weight dextran polymere. The dialysis membrane is permeable to all the relevant nutrients, but not to the dextran. The resulting osmotic pressure of the dextran is transmitted over the membrane onto the spheroid. It turns out that indeed pressures as low as $500Pa$ are able to significantly slow growth.

Unfortunately this approach is very time consuming, because of difficulties in handeling spheroids in dialysis bags. A simpler approach, is to use the outer cell membrane as dialysis membrane. At the concentrations necessary to exert pressures of the order of many kPa dextran does not effect cells if grown on substrates. One can thus assume it has no poisonous effect. Also, it can not penetrate into the spheroid. Thus, the main effect of dextran in the growth medium is to exert a well controlled mechanic stress on the growing spheroid. These experiments can be performed on a much larger number of spheroids, and significant statistics can be obtained. The results (See Fig. 10) confirm the growth reducing effect of pressure. However, some details come at a surprise. First of all, the tissue never shrinks due to pressure. Equation 1 would predict an exponential growth or shrinkage linearly depending on pressure. However, the growth

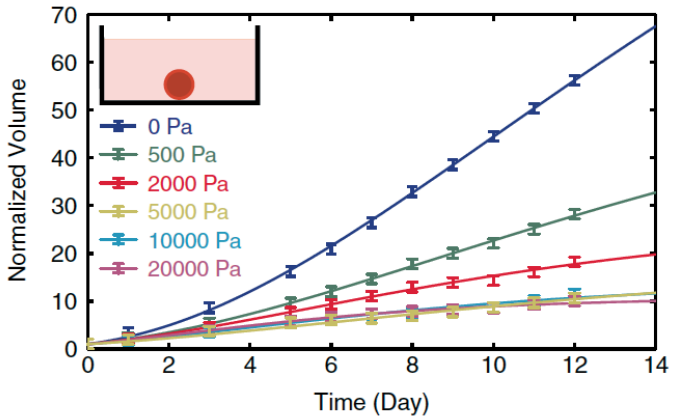


Fig. 10: *Growth of tissue spheroids under isotropic pressure. Cells from a colon carcinoma cell line (CT26) are placed in wells, and grown with standard growth medium with added dextran. The dextran exerts a well defined isotropic pressure on the spheroids. Volume is measured over the course of two weeks. From [15]*

reducing effect seems to saturate above 5kPa. Furthermore the growth is clearly not exponential. The answer is, that while pressure is constant across the system, the biochemical, but also the biomechanical, environment is not. While the first certainly has some effects, it is the latter that we focus on here.

Particle based simulations can proof the existence of the biomechanical differences in the environment. With the right parameters, the growth under pressure can be well reproduced. Because biochemistry is assumed uniform in the simulation, there is a mechanical explanation that accounts for the observed growth phenomena.

In order to increase volume, a cell has to deform its surrounding. I.e. it builds up a strain dipole. Insertion of a strain dipole in an elastic medium is much easier close to a free surface, than it is in bulk. Thus divisions at the surface are favored mechanically over divisions in bulk.

Indeed both simulations and experiments exhibit increased division close to the surface (See Fig. 11). Simulations allow for a much more detailed analysis of the division profile. The simulation data suggests that division is enhanced in a thin layer close to the interface, but than saturates below (!) the apoptotic rate. This finally explains the growth curves. On average, cells die more than they divide $(k = (k_d - k_a) < 0)$ but in a thin layer close to the surface, division is strongly increased (δk_s). Using simple cell number balance, we arrive at an equation for the growth of the spheroid:

$$\partial_t N = kN + N_s \delta k_s \tag{15}$$

with N_s the number of cells in the surface layer that is favored for division. Assuming a constant density, and a constant surface layer thickness λ, we get

$$\partial_t V = kV + \lambda \delta k_s (V^{2/3}(36\pi)^{1/3}) \tag{16}$$

Fitting Eq. 16 to the growth curves of simulations and experiments is able to reproduce the growth curves quite well, and allows to extract how the different growth rates depend on pressure. It turns out, that the surface rate $\lambda \delta k_s$ is largely unaffected, while k is negative for all applied pressures and decreases further with increasing pressure (See Fig. 12).

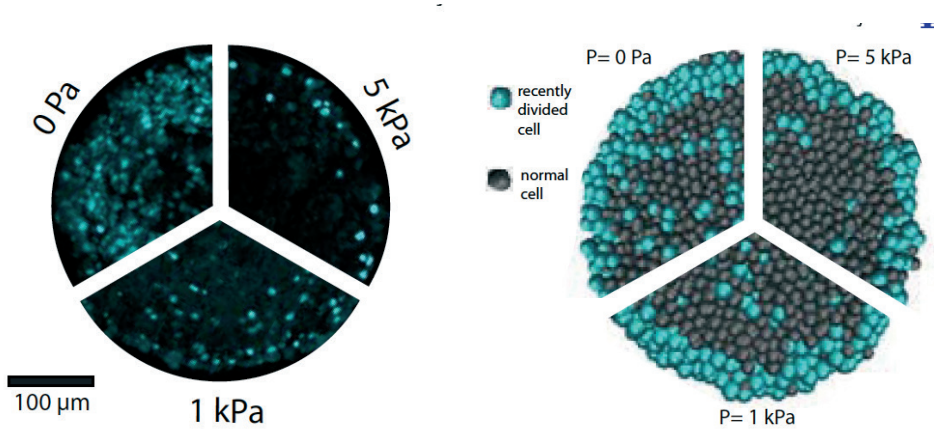


Fig. 11: *Cryosections and virtual cryosections of growing spheroids. In experiments, spheroids are sacrificed, frozen and cut into thin slices. By labeling Ki67 cell divisions are marked in cyan. In simulations, recently divided cells are marked in cyan, others in grey. Independent of pressure, most divisions happen at the surface, while the division is strongly suppressed by pressure in the bulk. From [16]*

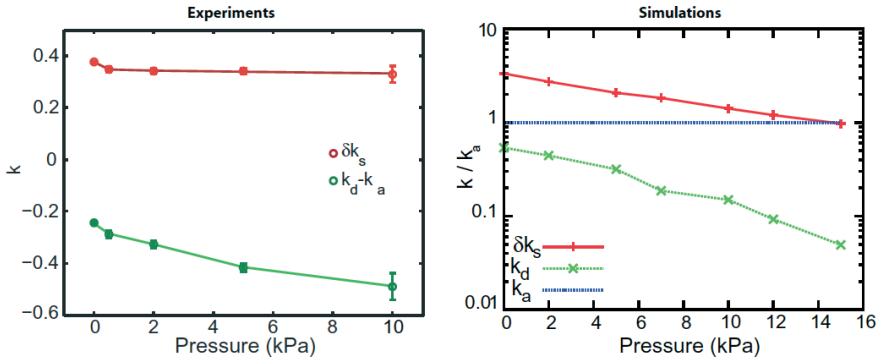


Fig. 12: *Growth rates extracted from growth curves by fitting eq 16. The growth rate at the surface δk_s is much less effected than the bulk growth rate k_d. From [16]*

The negative growth rate in the bulk of the spheroid, and division at the surface naturally lead to a stable steady state where a flow of cells from the surface balances the net death in the bulk. For an incompressible fluid of cells this flow can be estimated analytically. The conservation equation reads

$$\nabla \vec{v}(r) = k(r), \tag{17}$$

where we assumed radial symmetry. As above, the growth rate $k(r)$ is k in the bulk, and increased by δk_s within a range λ of the surface at position $R(t)$, i.e. $k(r) = k + \delta k_s \Theta(R(t) - r - \lambda)$ with Θ the heavyside step function. With the boundary condition of zero velocity at the center, we can solve for v and get:

$$v_r = \begin{cases} \frac{1}{3} k r & \text{if } r < R(t) - \lambda \\ \frac{1}{3}(k + \delta k_s) r - \frac{1}{3} \delta k_s \frac{(R(t)-\lambda)^3}{r^2} & \text{if } r > R(t) - \lambda \end{cases} \tag{18}$$

The spheroid grows with the velocity at the surface $\partial_t R(t) = v_r(R)$. Note that expanding $\partial_t R$ to first order leads back to equation (16). Thus both equations predict the same steady state size

$$R_\infty = -3\delta k_s \lambda / k \tag{19}$$

The flow of cells predicted by equation (18) can be measured experimentally by marking cells at the surface, and following them over the course of time[17]. This fit allows an independent measurement of the growth rates from the growth curves, and leads to very similar values.

7 Conclusions

Growing materials display a set of phenomena completely unknown to conventional materials. Key aspects are a self developed homeostatic stress, viscous behavior due to material turnover and competition. To understand the dynamics of growing materials continuum mechanics very similar to classic viscoelastic materials can be used. The main difference comes from an additional source (or sink) term in the continuity equation. Mesoscopic particle based simulations can bridge the difficult gap between analytic theory and real experiments, but more importantly provide a third perspective on the problems at hand.

The field of growing materials is still young, and many aspects have not been addressed yet. For example, both simulations and analytic model here, have a source term in the continuity equation, while of course in reality matter remains conserved. The truth is that nutrients and fluid (the "interstitial fluid" pass between the cells, and gets turned into tissue material. Thus the growth (and death) rate is indeed a conversion rate from interstitial fluid to tissue material. See for example Ref. [18] for a further discussion.

References

[1] D. P. Kiehart, C. G. Galbraith, K. A. Edwards, W. L. Rickoll, and R. A. Montague, J. Cell Biol. **149**(2), 471 (2000).

[2] A. E. Shyer, T. Tallinen, N. L. Nerurkar, Z. Wei, E. S. Gil, D. L. Kaplan, C. J. Tabin, and L. Mahadevan, Science **342**(6155), 212 (2013).

[3] W. Walshe, *The Nature and Treatment of Cancer* (Taylor and Walton, 1846).

[4] A. W. Orr, B. P. Helmke, B. R. Blackman, and M. A. Schwartz, Dev. Cell **10**(1), 11 (2006).

[5] M. A. Wozniak and C. S. Chen, Nat. Rev. Microbiol. **10**(1), 34 (2009).

[6] C.-P. Heisenberg and Y. Bellache, Cell **153**(5), 948 (2013).

[7] E. Hannezo, J. Prost, and J. . F. Joanny, Phys. Rev. Lett. **107**(7), 078104 (2011).

[8] S. Mukherjee, *The Emperor of All Maladies* (Scribner, 2010).

[9] D. T. Butcher, T. Alliston, and V. M. Weaver, Nat. Rev. Cancer **9**(2), 108 (2009).

[10] A. Fritsch, M. Hockel, T. Kiessling, K. D. Nnetu, F. Wetzel, M. Zink, and J. A. Käs, Nature Physics **6**(10), 730 (2010).

[11] O. J. T. McCarty, M. R. King, and P. A. Insel, American Journal of Physiology - Cell Physiology **306**(2), C77 (2014).

[12] L. Landau, E. Lifshitz, A. Kosevich, and L. Pitaevskiĭ, *Theory of Elasticity* (Butterworth-Heinemann, 1986).

[13] L. Landau and E. Lifshitz, *Fluid Mechanics* (Elsevier Science, 2013).

[14] M. Basan, T. Risler, J. Joanny, X. SastreGarau, and J. Prost, HFSP Journal **3**(4), 265 (2009).

[15] F. Montel, M. Delarue, J. Elgeti, L. Malaquin, M. Basan, T. Risler, B. Cabane, D. Vignjevic, J. Prost, G. Cappello, and J.-F. m. c. Joanny, Phys. Rev. Lett. **107**, 188102 (2011).

[16] F. Montel, M. Delarue, J. Elgeti, D. Vignjevic, G. Cappello, and J. Prost, New Journal of Physics **14**(5), 055008 (2012).

[17] M. Delarue, F. Montel, O. Caen, J. Elgeti, J.-M. Siaugue, D. Vignjevic, J. Prost, J.-F. m. c. Joanny, and G. Cappello, Phys. Rev. Lett. **110**, 138103 (2013).

[18] J. Ranft, J. Prost, F. Jülicher, and J.-F. Joanny, Eur Phys J E Soft Matter **35**(6), 46 (2012).

E 2 Protein Folding: Kinetics, Pathways, Landscapes

J. Fitter

Institute of Complex Systems / Molecular Biophysics (ICS-5),

Forschungszentrum Jülich GmbH and

I. Physikalisches Institut (IA), AG Biophysik, RWTH Aachen

Contents

1 Introduction **2**

2 Folding Pathways **2**
2.1 Mechanisms of Protein Folding 2
2.2 Thermodynamic Hypothesis and Free-Energy Landscapes 4

3 Monitoring Folding Pathways in Time and in Space **5**
3.1 Folding Kinetics and Related Time Scales 6
3.2 Multi-Pathways and State Connectivity 8

4 Coupling of Inter-molecular Binding and Folding **10**

5 Chaperone-assisted Folding **12**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

It is still one of the great unsolved problems in life science that we do not understand in detail how an amino acid sequence of a polypeptide chain is transformed into a well defined three-dimensional protein structure (also called folding problem). This fundamental process and the attempts to establish a more complete understanding of the so called protein folding, is a prominent topic in the field of biophysical chemistry and biochemistry. In addition the analysis of protein folding and stability has also become more and more relevant for medicine and biotechnology. There are at least three major reasons that an increasing theoretical and experimental interest in protein folding exists [1, 2]: (i) Due to efforts on genome sequencing projects the acquisition of DNA sequences is increasingly faster nowadays. Compared to that, the acquisition of three-dimensional proteins structures is still slow, limited by the time consuming process of searching for proper crystallization conditions. In this respect the knowledge how the linear sequence of amino acids is translated into spatial information is the missing link. (ii) There is a tremendous interest in the over-expression of recombinant proteins for industrial, biotechnological, and research applications. (iii) Incorrect folding or misfolding of proteins is often related to protein aggregation and fibrillogenesis, which is connected to a number of serious diseases, such as BSE, or Huntingtons and Alzheimer diseases. Essentially the folding problem can be subdivided in three basic problems [3]. First, there is the question of the folding code. What are the forces that dictate the protein structure for a given amino acid sequence? Second, there is the challenge of a protein structure prediction. How can we predict a native three-dimensional protein structure from its amino acid sequence? And a third aspect deals with the understanding of the folding process itself. How are the routes or pathways during protein folding and why does the folding happens in some cases very fast and in others much slower (folding kinetics)? All these aspects are still intensively under investigation which is also supported by the fact that the number of research reports and review articles on various aspects of protein folding has increased dramatically over the last decades [4].
Herein we discus the state-of-the-art concepts and models in protein folding with an emphasis on single-molecules studies which give access to a detailed characteristics of the folding pathways. Another focus deals with inter-molecular binding of various interaction partners, like ions, substrates, or macromolecular ligands which can have a profund impact on the properties of the underlying folding process. Finally the folding of proteins under *in vivo* conditions is discussed which is supported by cellular chaperones helping proteins in the cell to achieve more efficiently to reach the native state.

2 Folding Pathways

2.1 Mechanisms of Protein Folding

A variety of studies have shown that the folding of many small proteins can be characterized by considering solely two states, where only fully unfolded and completely folded states are occupied with any significant population. In theses cases there is no buildup of intermediated states as folding occurs. However, most proteins, particularly larger ones, do not exhibit such a simple folding behaviour. Instead, folding involves intermediate states, which are partially folded structures that are stable, although only transiently. Various models illustrate schematically how different folding events occur sequentially in a so-called folding pathway (Fig. 1). Since na-

tive proteins usually contain substantial amounts of secondary structure, it seems reasonable to assume that such structures would form early in the folding process.

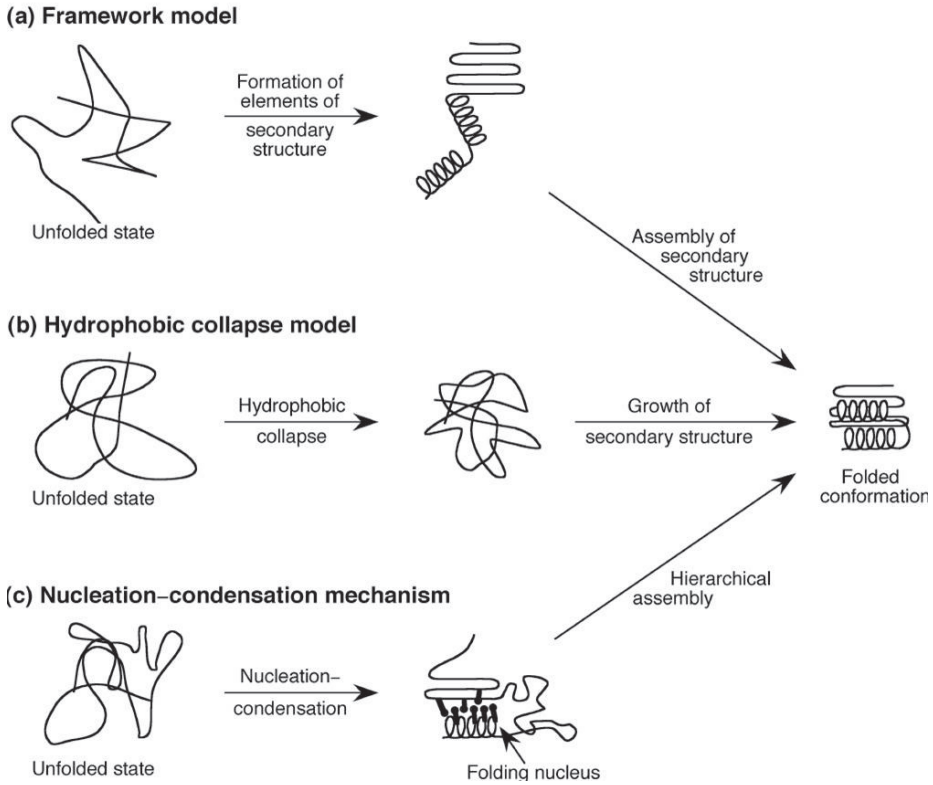


Fig. 1: *Models for protein folding. (a) In the framework model protein folding is thought to start with the formation of elements of secondary structure. These elements form independently of tertiary structure, or at least before tertiary structure is locked in place. The elements then assemble into the tightly packed native tertiary structure by diffusion and collision. (b) In the hydrophobic collapse model for folding the initial event of the folding reaction is thought to be a relatively uniform collapse of the protein molecule, mainly driven by the hydrophobic effect, i.e., the tendency of non-polar groups dissolved in water to cluster together. Stable secondary structure elements can only form in the resulting collapsed state. (c) In the nucleation-condensation mechanism early formation of a folding nucleus catalyzes further folding. The nucleus is diffuse, but comprises secondary structure interactions and approximately correct tertiary structure interactions. This model is consistent with the funnel model (see Section 2.2) which focuses on the rapid decrease of the conformational dispersity in the course of the reaction. Taken from [2]*

Indeed, various studies showed that local structures, like helices or harpins, form fast relative to the time of the complete folding process (framework or diffusion collision model). However, from other studies we learned that various distant segements of the polypetide chain come together without forming secondary structure, stabilized instead by sidechain interactions (nucleation-condensation mechanism). Apparently, the order in which secondary and tertiary

interactions are formed can vary from protein to protein [5]. There are at least two facts related to the ”folding problem” for which the existence of intermediate states is of specific importance. (i) In the late sixties of the previous century Cyrus Leventhal pointed out that protein folding cannot be explained by an exhaustive search of conformational space. The number of possible conformations is so large that is would take more than the age of the universe to sample them all (Leventhal‘s paradox). Hence protein folding must occur by some directed process, for example by forming intermediate states which makes protein folding a hierarchical process. (ii) The unfolded state is characterized by a pronounced structural flexibility resulting in a heterogeneity of conformational states present in an ensemble of molecules in the sample volume. As a consequence folding routes of individual proteins can be quite different in the folding process caused by different ”starting structures of the unfolded state”. Therefore the application of single molecule techniques is of particular importance, since these techniques can unravel details of the folding pathways which are hidden in data obtained from bulk measurements (see Section 3 and 4).

2.2 Thermodynamic Hypothesis and Free-Energy Landscapes

The idea that polypetide chains include already all of the information necessary to fold and that the folding process itself is thermodynamically driven is known as the thermodynamic hypothesis. In this respect the thermodynamic hypothesis states that the native structure of a protein corresponds to a minimum in the free energy of a protein (Fig. 2).

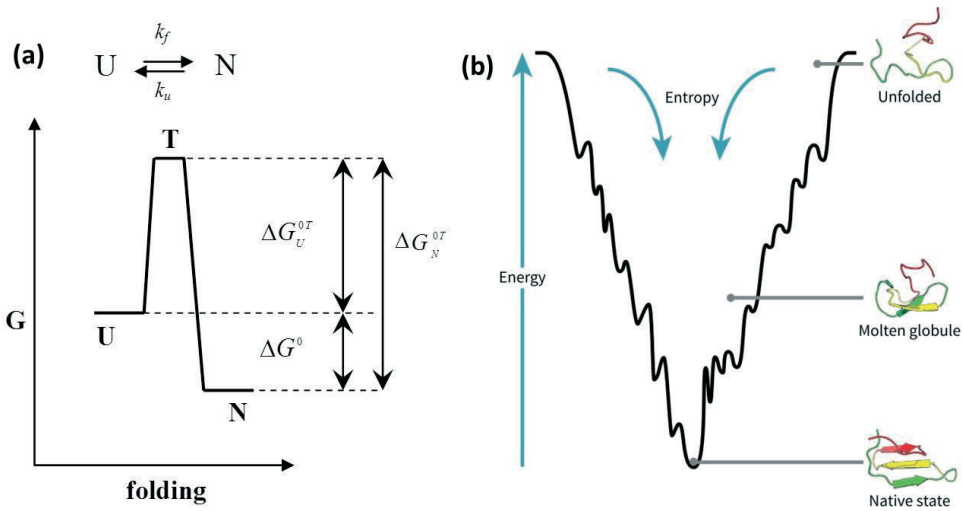


Fig. 2: *(a) Free energy profile for a two-state reaction where the native N and the unfolded U state are separated by a free energy barrier (transition state T). (b) Schematic presentation of a folding energy landscape (folding funnel). Unfolded states of a protein are characterized by a higher level of free energy (shown at the top of the funnel). Some of the proteins are trapped transiently in intermediate states (here shown as molten globules) before they reach the one and only global minimum of the native state.*

For the simple case of a reversible two-state process (see above) transition state theory allows a

straightforward description relating the difference in free energy ΔG^0

$$\Delta G^0 = G^U - G^N = \Delta H - T\Delta S \quad (1)$$

to folding and unfolding rates, k_f and k_u, respectively (Fig. 2a). As shown in equation 1, the protein stabilization is determined by entropic as well as by enthalpic contributions. While enthalpic contributions often play a role through intermolecular bonds (H-bonds, S-S-bonds, salt bridges) the entropy change upon folding is dominated by the conformational freedom of the protein structure and by the interaction of the protein with hydration water. A protein structure is controlled by a subtle balance of stabilizing and destabilizing forces, and the resulting energy (i.e. ΔG^0) which shifts the equilibrium to the native state (under physiological conditions) is often not larger than 1-4 kJ/mol (i.e. a few RT at 25° C). All the above mentioned considerations hold for the case of reversible processes. The height of the transition state barrier (T) determines how fast the system will reach equilibrium.
The folding of a protein involves formation and breakage of a multitude of intra-molecular contacts. Due to this fact a single-trajectory view (i.e. one more or less well defined route from the unfolded to the native state) of protein folding is an over-simplification of the folding process at a molecular level. A more realistic view is given by a perception of protein folding that is characterized by a myriad of many different folding routes. Such a picture emerges from a multidimensional rugged energy landscape with a funnel like shape (see Fig. 2b) in which individual molecules find different routes from the unfolded to the native state. For proteins that would fold without observable intermediates (i.e. long- lived states) the energy surface would be rather smooth. The relevance of intermediates in the folding process is still in debate. In some cases intermediate states represent important structures with a high content of native contacts which are assumed to be productive to reach the final native state (on-pathway intermediates). In other cases mainly non-native interactions are trapped and therefore theses states are called off-pathway intermediates. However, also off-pathway intermediates are of interest since they reflect properties of the energy landscape and might represent substrates for chaperones (folding helper proteins) or play a role for protein aggregation in the cellular context (see Section 5).

3 Monitoring Folding Pathways in Time and in Space

Experimentally properties of the folding process are often studied by pushing the equillibrium between the native and the non-native state(s) (or unfolded state), which favors under physiological conditions the native state (see Fig. 2), towards the non-native states. In this manner unfolded states (as well as intermediate states) are getting more and more populated and the time course of the transitions and the structural as well as dynamical properties of non-native states can be characterized in detail. For this purpose in most cases increased temperatures (i.e. thermal unfolding) or solutions with a high concentrations of chemical denaturants, like urea or guanidine hydrochlorid (i.e. chemical unfolding), are employed. By varying these environmental parameters unfolding as well as folding transitions can be induced. In addition by keeping the sample at a intermediate value (mid concentration of the chemical denaturant or melting temperature) we will obtain permanent and repetitive transitions of individual molecules between two states, an approach often employed in single molecule studies.

3.1 Folding Kinetics and Related Time Scales

The knowledge about the kinetics of the folding process is extremely valuable for the understanding of the energy landscape, for example in terms of local saddle points (transition states) or of local minima (intermediates). Assuming a simple two-state process the free energy barrier separating the native and the unfolded state is essentially rate limiting the process. The unfolding and the folding rate constants, k_u and k_f, are given by:

$$k_u = k_0 \cdot e^{\frac{-\Delta G_N^{0T}}{RT}}, k_f = k_0 \cdot e^{\frac{-\Delta G_U^{0T}}{RT}} \tag{2}$$

In this equation k_0 represents the pre-exponential factor and $\Delta G_N^{0T}$the free energy of the unfolding barrier (see Figure 2). Examples of unfolding kinetics, induced by chemical denaturants follow a mono-exponential behavior (see equation 2). In a similar manner the folding rate constant k_f is related to the free energy barrier height ΔG_U^{0T}, as shown in Figure 3 (right upper panel). Assuming a reversible process we obtain an equilibrium constant

$$K_{eq} = \frac{[N]_{eq}}{[U]_{eq}} = \frac{k_f}{k_u} \tag{3}$$

wich is related to the free energy ΔG^0 stabilizing the native state under given conditions:

$$\Delta G^0 = -RTln \cdot K_{eq} = -RT \cdot ln(\frac{k_f}{k_u}) \tag{4}$$

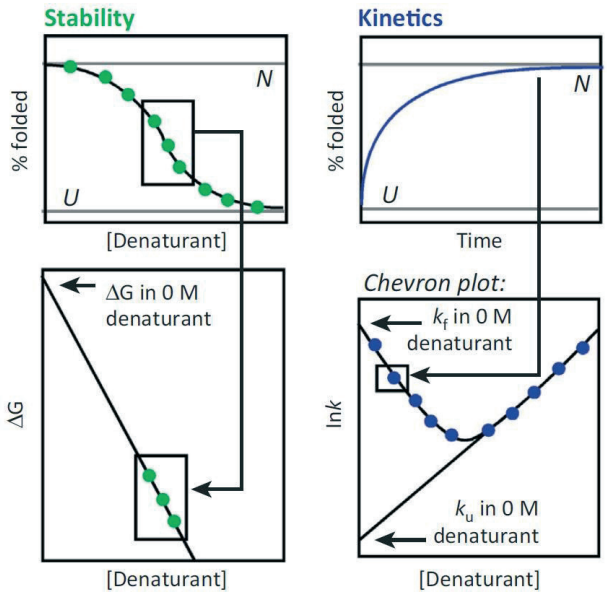


Fig. 3: *For proteins that do refold reversibly, there are widespread assays available to study protein stability and folding kinetics. For such assays, conditions must be identified where folding is reversible. Under these conditions, the free energy of folding (ΔG^0) can be determined either from equilibrium denaturation measurements (left; green) or from kinetic measurements of rate constants (k_f and k_u) (right; blue). Taken from [6].*

The experimental determination of ΔG^0 is often accomplished by measuring the fraction of folded and unfolded protein ($[N]$, $[U]$) or by measuring the corresponding transition rates (k_u, k_f) as a function of the denaturant concentration $[D]$. It was found emperically that the obtained ΔG^0 depends linearly on $[D]$ with

$$\Delta G^0 = \Delta G^0(H_2O) + m_{eq} \cdot [D] \tag{5}$$

Here m_{eq} describes the difference in the solvent accessibility to the surface between the folded and the unfolded state. As demonstrated in a so-called Chevron-plot (lower right panel in Fig. 3) the dependence of k_u and k_f on the concentration of GndHCl is described by the following relation:

$$k_u = k_u^w \cdot e^{m_u \cdot [D]}, k_f = k_f^w \cdot e^{m_f \cdot [D]} \tag{6}$$

with $m_{eq} = m_u + m_f$. Furthermore, k_u^w and k_f^w are the corresponding rates for unfolding and folding in the absence of denaturant.
The discussed models of transition state theory were introduced by Eyring for simple reactions in the gas phase [7] and for these cases the pre-exponential factor (i.e. a fundamental rate constant, see equation 2) was given by

$$k_0 = \kappa \frac{k_B T}{h} \approx 6 \cdot 10^{12} s^{-1} \tag{7}$$

at 25°C. Here κ is the transmission factor (typically assumed to be one), k_B is the Boltzmann constant, h is the Planck constant, and T the absolute temperature. In contrast to assumptions made in the Eyring theory, protein folding/unfolding transitions involve formation and breakage of many weak intramolecular interactions rather than a formation or cleavage of a single covalent bond. Therefore the above given value of k_0 will not be useful for our kind of analysis. For protein folding/unfolding reactions segmental diffusion processes of polypeptide chain fragments and internal friction will determine the rate limitation. In principle this factor can be quite different for individual proteins. Most probably it will also depend on the polypeptide chain length (i.e. on the proteins size). As an improvement in this respect, Kramers developed a theory on how chemical reaction rates are influenced by the viscosity of the medium. Kramers′ theory assumes that the dynamics can be described by one-dimensional diffusion along a reaction coordinate in which both the reactant wells and the barrier top are parabolic (Figure 4A). The corresponding (folding) rate constant is given by:

$$k_f = \frac{2\pi k_B T}{\omega_{min} \omega_{max} D_{max}} \cdot e^{\frac{-\Delta G_U^0}{RT}} \tag{8}$$

where ω_{min} and ω_{max} are frequencies that characterize the curvature of the free energy profile at the unfolded well and (inverted) barrier top, respectively, and D_{max} is the diffusion constant at the barrier top. As shown by Monte Carlo simulations the accuracy of equation 8 is not declined if we assume $\omega_{min} = \omega_{max}$ and $D_{min} = D_{max}$. This approximation leads to

$$k_f = \frac{2\pi k_B T}{\omega_{min}^2 D_{max}} \cdot e^{\frac{-\Delta G_U^0}{RT}} \tag{9}$$

and can be used to obtain the pre-exponential factor and therefore to determine the speed limit of the folding process (for more details see [8]).

Eaton and co-workers showed recently that the Kramers diffusion coefficient and the free-energy barrier can be characterized by measuring the temperature- and viscosity-dependence of the transition path time for protein folding [9]. The transition path is the small fraction of an equilibrium trajectory for a single molecule when the free-energy barrier separating two states is actually crossed (see shaded area in Fig. 4A). Its duration, the transition path time, can now be determined from single protein molecules undergoing folding/unfolding transitions (Fig. 4B).

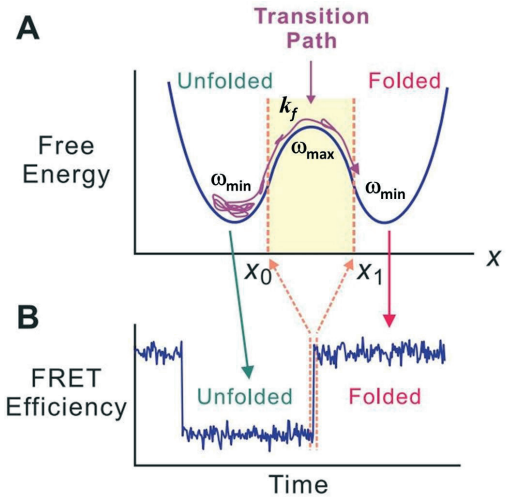


Fig. 4: *Folding transition path for a two-state protein. (A) The kinetics of protein folding is described by energy landscape theory as diffusion on a one-dimensional free-energy surface with an order parameter (x) as a reaction coordinate. (B) FRET efficiency trajectory. In the typical experiment, the donor and acceptor FRET fluorophores are attached to cysteine residues, which are closer on average in the folded state (higher FRET efficiency) than in the unfolded state (lower FRET efficiency). Taken from [9].*

In particular time resolved single molecule Förster resonance energy transfer (FRET) studies can reveal valuable information in this respect (for methodical details see for example Chapter C3, Single-Molecule Fluorescence Spectroscopy). As shown in Figure 4B it is quite straightforward to obtain the transition rate constants (k_u and k_f) from single molecule trajectories, which measures the frequency how often a transition occurs. In contrast it is much more elaborative to determine the transition-path time which is the duration of a successful barrier-crossing event. However, successful studies revealed that for two small proteins which differ by four orders of magnitude in k_f ($10^4 s^{-1}$ and $1 s^{-1}$) the corresponding transition path-times differ by less than five-fold (2 μs and $< 10\mu$s). This interesting result supports model predictions which claim that the barrier height is not affecting the transition path-times. In fact the latter are strongly determined by the roughness of the underlying energy landscape which is closely related to the internal friction of the polypetide chain [9].

3.2 Multi-Pathways and State Connectivity

For a long time it was assumed that proteins fold through distinct intermediates and in predetermined pathways. By establishing funnel-shaped energy landscapes another more statistical

picture of protein folding emerged which assumes multiple unpredictable folding routes and intermediate conformations. In particular the folding of larger multi-domain proteins can occur in a complex network of on and off-pathway intermediates. Recent theoretical as well as experimental methodological advances made it possible to given answers to the question whether proteins fold through a limited number of distinct obligatory intermediate states in an ordered kinetic sequence or through a more heterogeneous collection of independent multiple routes [10].

The most impressive results in this respect are coming again from single molecule studies. As a major technique in single molecule studies, optical traps or tweezers are ideally suited to measure rather weak forces (below pN) which allows for a detailed study of protein folding pathways. By using optical tweezers experiments on full length calmodulin it was shown that the folding pathway involves at least 6 states connected by 6 transition pathways [11]. In so-called constant force experiments both the folded and the unfolded states are populated and protein folding/unfolding transitions occur, which are measured by abrupt length changes (Fig. 5A,B).

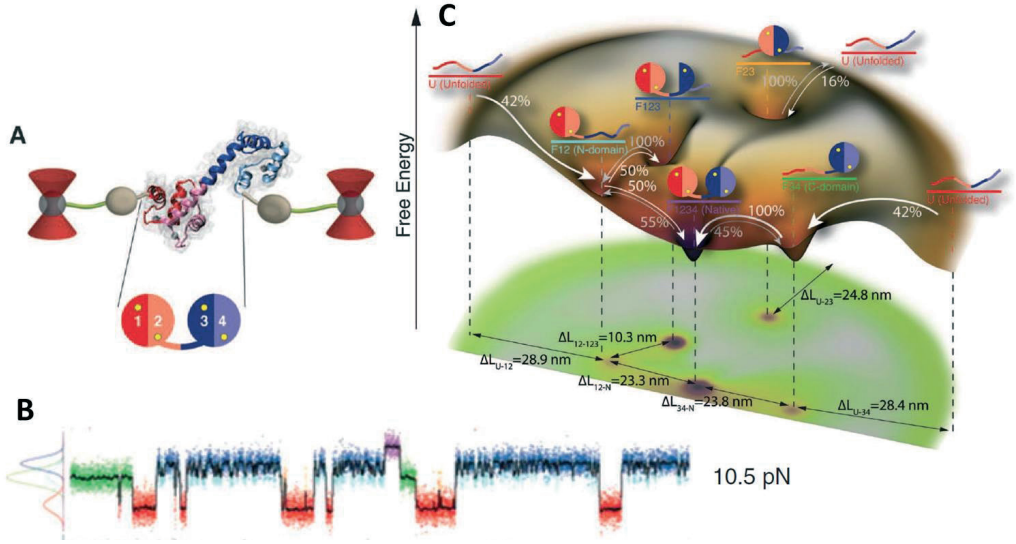


Fig. 5: *(A) Schematic of a dual-trap setup for pulling experiments using optical tweezers. Beads trapped by the foci of two laser beams bound to DNA handles which are connected to the protein can be used to exerted a pull on the protein structure. By means of this setup lengths in the nanometer regime and pN forces can be measured. (B) Example of a measured trace performed with a single molecule held at constant load over a long period of time and equilibrium fluctuations in extension reflecting folding/unfolding transitions can be observed in a real-time. (C) Full kinetic network of calmodulin folding and unfolding. Arrows show all observed transitions. The percentage values provided for each transition give the fraction of transitions along the respective pathways out of each state. Distances in the lower part are differences in contour length. Taken from [11].*

The resulting occupation probabilities and the dwell time distributions of indivudal states can be used to calculate force-dependent microscopic kinetic rate constants and equilibrium free en-

ergies. For calmodulin this approach revelad a detailed picture (Fig. 5C) with two off pathway intermediates (F23, F123) which exhibit non-native interdomain interactions and compete with the very fast productive folding pathway (via F12 and F34).
In another example, employing single molecule FRET, the folding landscape of adenylate kinase was investigated. In this study single molecule trajectories (see Fig. 6B) were measured with surface tethered proteins which were kept at different concentrations of a chemical denaturant, ensuring to populate native as well as non-native states. By using hidden Markov modelling six metastable intermediates were identified [12]. Adenylate kinase consists of three domains strongly interacting with each other (Fig. 6A). Therefore the individual domains cannot be seen as individual folding units, which is reflected in a complexity of the folding dynamics already reported in some earlier studies. This complex behaviour was unveiled in more detail by an extensive analysis of thousands of single molecule trajectories. The important outcome of the study is given by the fact that the connectivity of intermediate states depends on the denaturant concentration, where only at low concentrations multiple intersecting folding pathways co-exist.

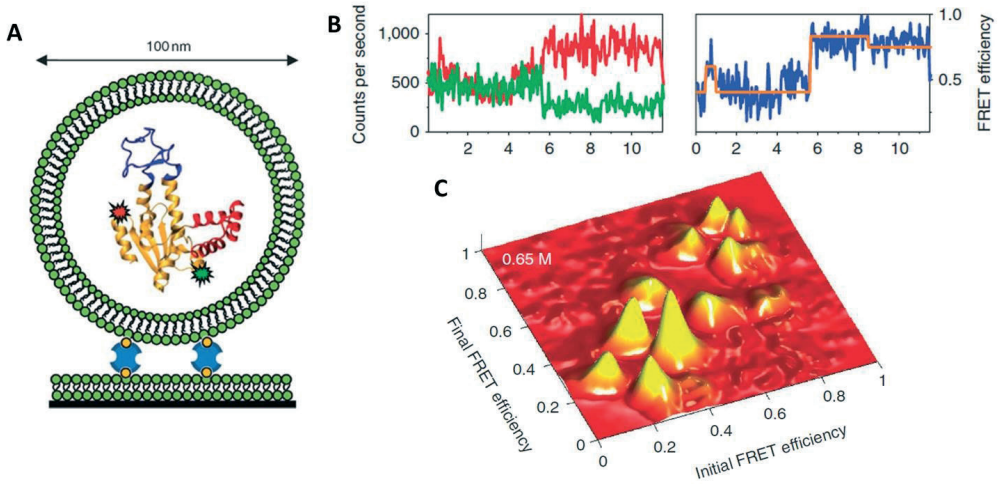


Fig. 6: *(A) An individual adenylate kinase molecule (not drawn to scale with the vesicle), double-labelled for FRET, is encapsulated in vesicles tethered to a glass-supported bilayer using biotin-streptavidin chemistry. (B) Example of a fluorescence trajectory of individual proteins, showing transitions between three different FRET efficiency levels. (C) Example of a transition map at 0.65 M GndHCl concentration, constructed from the experimental data using hidden Markov modelling analysis results. Taken from [12].*

With increasing denaturant concentration the folding landscape becomes more and more sequential. Additional complexity in the folding process shows up if the folding reaction takes places in the presence of ligands, as we will see in the next Section.

4 Coupling of Inter-molecular Binding and Folding

As known from various studies different types of ligands can have an impact on protein folding. In this respect small ligands like metal ions are known for a long time to be part of the native

structure and are often even required to reach the native state (see Fig. 7A).

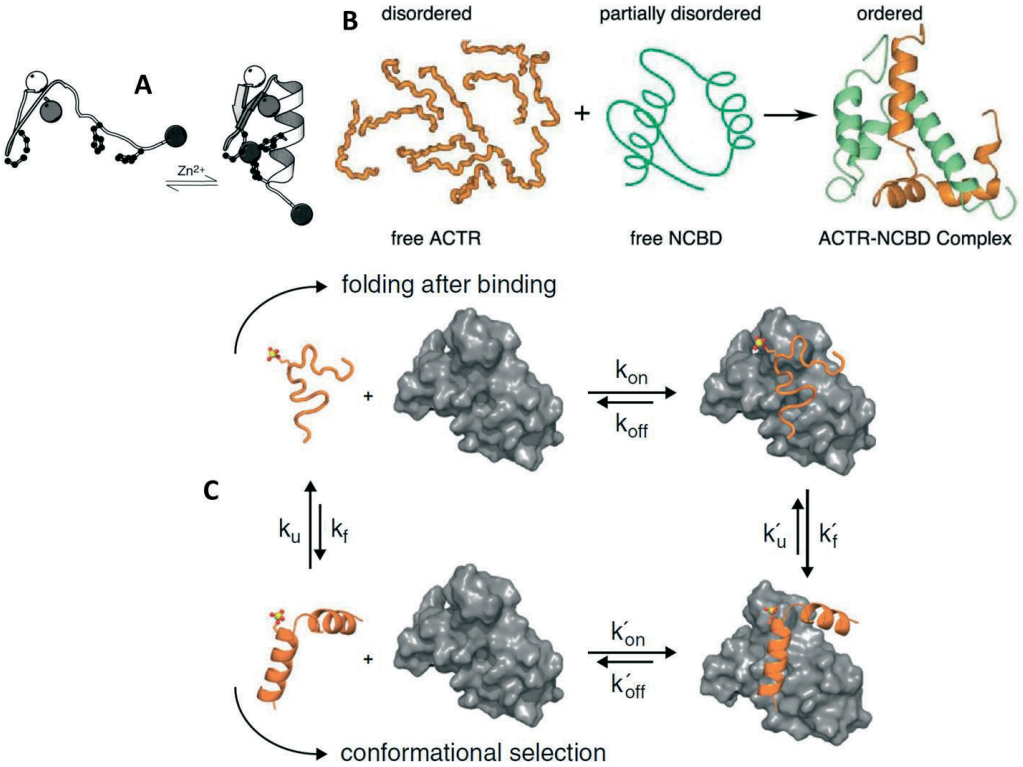


Fig. 7: *Examples of coupling between binding and protein folding: (A) Zn-finger peptide undergoes folding transition upon metal binding. (B) Coupled binding and folding of ACTR of p160 nuclear hormone receptor and NCBD of CBP/p300 transcriptional coactivator (for details see [14]). (C) Schemes of the folding after binding (similar to induced fit model) and of the conformational selection mechanism. Taken from [15].*

In the recent past numerous proteins have been identified which are to large degree disordered and show a transition to a more odered (native) stucture mainly in conjunction with binding partner interactions. The understanding of the functional role in biology of theses intrinsically disordered regions in proteins or intrinsically disordered proteins (IDRs and IDPs, respectively) is a relatively new field in structural biology and requires a new perspective and new tools for successful investigations [13]. Often both interaction partners are proteins which both exhibit disordered structures if they reside isolated, but associate with a high affinity and form cooperatively folded structures (see example Fig. 7B). Key questions in this topic are for example whether folding or binding occurs first or whether the reaction mechanism involves multiple states and folding intermediates.

In order to characterize the underlying mechanisms of coupled binding and folding reactions time-resolved experimental approaches are required. In principle two extreme mechanisms describing the sequence of reactions are given by kinetic mechansims shown in Figure 7C. In the folding-after-binding mechanism binding to the ligand occurs to an unfolded or partially folded state and folding occurs subsequently in the complex. In the conformational selection mecha-

nism the low populated folded state selectively binds to the ligand, which results effectively in the stabilization of the native state [15].
A distinction between the mechanisms is possible by various experimental approaches employing time resolved optical spectroscopy. Here the most straightforward method is to induce folding by mixing the unfolded protein with its ligand and to monitor the reaction. In addition NMR relaxation dispersion experiments provide an alternative method to discriminated between both reaction mechanisms. Most experimental studies indicate that typically unfolded or partially folded states interact with the ligand and that the major folding barrier is generally crossed when the protein is bound to the ligand. This observation, supporting the folding after binding or induced fit model, was also made for various IDPs. Besides NMR relaxation dynamics and MD simulations, also single molecule FRET studies were successfully employed not only to resolve binding and folding reations, but also to characterize the structural flexibility of the various states occuring on the folding pathway (see for example [16]). In particular for IDPs the observed ligand induced order-disorder transition is believed to play a crucial role in the functional manifoldness of these proteins. Related to the fact that IDRs can interact with multiple binding partners the idea of binding-modulated functions of the IDPs emerged and specific classes of proteins, such like hub proteins or scaffold proteins, were already identified to operate in this manner (see [13] and references therein).

5 Chaperone-assisted Folding

During the vectorial appearance of newly synthesized polypeptide chains at the tunnel exit of the ribosome, hydrophobic residues are exposed to the cytosol. Because folding occurs in a highly crowded environment (300 - 400 $g\ l^{-1}$), these exposed residues can interact with a multitude of cellular components before they gain their native structure. These undesired interactions often irreversibly impair the folding process (see Fig. 8A).
In contrast to the relative fast formation of secondary structures (on the microsecond time scale) folding into the tertiary structure may need up to minutes. Therfore a high risk of misfolding and of aggreation exists, which can also lead the formation of amyloid fibrils which are assumed be toxic to cells and give rise to some neurodegenerative diseases. In order to lower the risk of pathological interactions, the cell is equipped with numerous types of molecular chaperones [18, 19]. For this purpose a network of cooperating chaperones, which in *E. coli* include the Trigger factor (TF), the ATP-dependent DnaK, and GroEL/GroES systems, assist in de novo folding (Fig. 8A). ATP-independent chaperones such as the TF or small heat shock proteins are believed to suppress the aggregation of exposed polypeptide chains and to delay the folding. They are the only chaperones known to utilize a ribosome docking site for nascent chain handling. Small nascent chains ($\sim$ 70 % of the total number of proteins) probably interact generally with the TF and may fold rapidly during synthesis (i.e. co-translational folding) without further assistance. Longer chains interact subsequently with the DnaK/DnaJ system and fold upon one or several cycles of ATP-dependent binding and release. Finally the GroEL/GroES system acts posttranslationally promoting the last assisted step in protein folding. The degree by which chaperones are needed varies from protein to protein. It is assumed that for small single domain proteins the engagement of chaperones is less crucial as compared to larger multi domain proteins. Furthermore the above mentioned systems for assisted folding exhibit functional redundancy which is also related to the fact that TF and DnaK are dispensible at moderate temperatures. Interestingly, at temperartures above 30 °C cells missing TF and DnaK exhibit

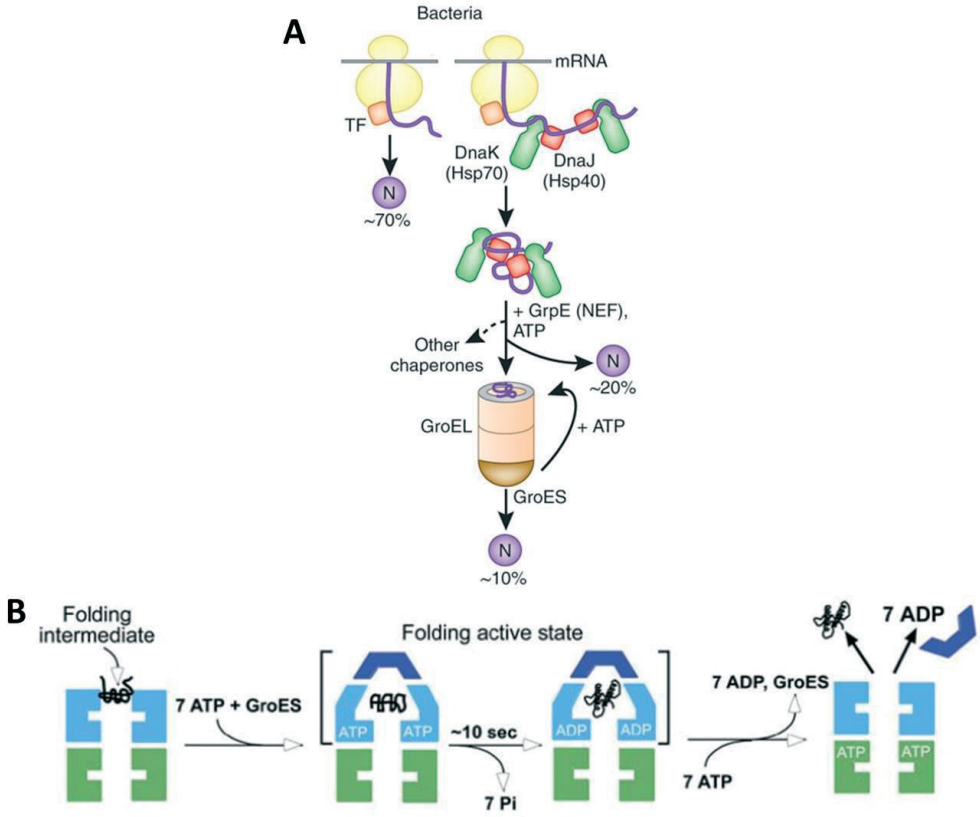


Fig. 8: *(A) Scheme for chaperone-assisted folding of newly synthesized polypeptides in the cytosol of bacteria. (B) Chaperone mechanism in promoting folding through kinetic partitioning. Reaction cycle of the GroEL/GroES system with switching between high- and low-affinity states for unfolded and partially folded protein by ATP binding and hydrolysis. Taken from [17].*

massive aggregation which can be compensated by the overproduction of GroEL/GroES (for details see [19] and references therein).

As mentioned above GroEL acts protstranslationally and recieves its substrate (a uncompletely folded polypeptide chain) through transfer from upstream chaperones, here the DnaK/DnaJ chaperonin. The basic mechanism of GroEL and its cofactor GroES involves encapsulation of a single substrate molecule in a cage-like structure (Fig. 8B). Here the double ring of GroEL 14mers and the GroES 7mer which act as a ATP controlled lid ensure a reversible closing and opening of the cage. In this so called "Anfinsen cage" GroEL catalyzes the rescue of kinetically trapped states of the protein and provides a confined folding space in an environment free of aggregation. By this the environment of the cage appears to modulate the energy landscape and the trajectories along which folding proceeds, resulting in an accelerated folding process. Interestingly, substrates can shuttle bidirectionally between DnaK and GroEL indicating that these interconnections setup a flexible network of chaperones.

Understanding the mechanisms underlying the process of folding *in vivo*, requires information about the substrate (protein) conformations populated along the chaperone assisted folding path-

way. In a single molecule FRET study a specific mutant of the maltose binding protein (double mutant V8G and Y283D) was studied with respect to its folding properties (Fig. 9). This mutant (DM-MBP) is an ideal substrate for investigating how GroEL/GroES systems accelerate protein folding, as the lone DM-MBP folds spontaneously over several minutes. In contrast, chaperone assisted folding is 13 times faster [20]. This study shed light on details how GroEL and GroES determine the progession of folding. While the denatured state exhibit an expanded structure (having a low FRET effieciency with $f_E \sim 0.1$) the sucessfully refolded structures (spontaneously and chaperone assisted) show a rather compact structure ($f_E \sim 0.85$), as shown in Figure 9. The important intermediate state, where DM-MBP ist still bound to GroEL (in the absecnce of GroES) shows a heterogeneous conformational distribution of substrate molecules. Obviously the interaction of the substrate with GroEL induces substrate molecules populating both compact ($f_E \sim 0.66$) and at least locally expanded (unfolded with $f_E \sim 0.06$) structures. Subsequent ATP binding to GroEL initates GroES bonding and release of the substrate to inner volume of the folding cavity (see also Fig. 8B). By further ATP binding the substrate is released upon GroES dissociation. If the released substrate is still incompletely folded a new cycle can be started. Here the underlying concept is to enclose unfolded or only partly folded proteins, one molcecule at the time, in a specialized folding compartment which provides an effective solution to overcome the aggregation problem.

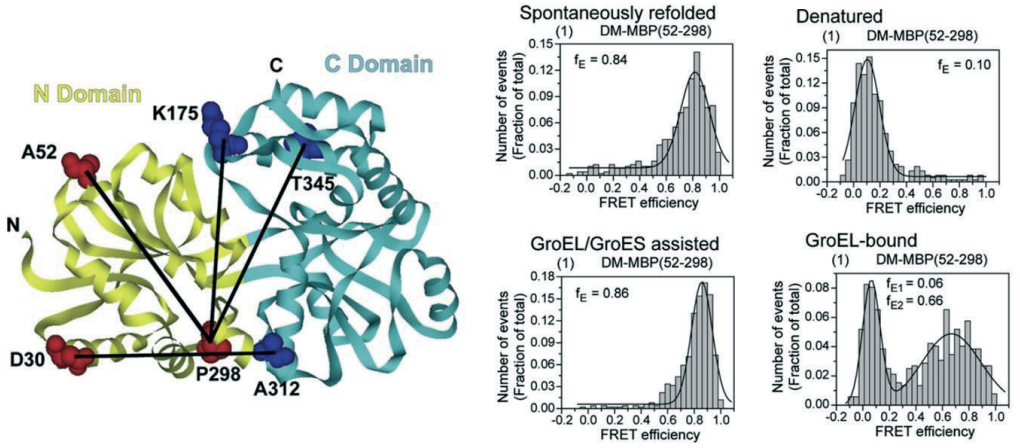


Fig. 9: *On the left side both domains of the maltose-binding protein (MBP) are shown together with positions of engineered cysteines used for dye attachments in single molecule FRET studies. On the right side single molecule FRET histograms are shown for MBP labeled at positions 52 and 298. Chemically denatured protein was diluted either in buffer alone (Spontaneously refolded) or into a buffer containing the full chaperonin system (GroEL/GroES assisted) to allow refolding in both cases. Furthermore data is shown for samples kept unfolded at 3 M GndHCl (Denatured) or which were transfered in buffer containing GroEL alone (GroEL-bound). Taken from [20].*

A further key aspect in resolving the mechanism of GroEL/GroES assisted protein folding is the conformational dynamics of the chaperonin system itself. In this respect several studies were performed which focussed not only on conformational changes of the GroEL complex but also on GroEL-GroES binding and unbinding during the working cycle (for details see [18] and references therein). In addition to fluorescence based single molecule techniques also force based

single molcecule methods, like atomic force microscopy or optical tweezers contribute already significantly to understand chaperone assisted folding [21].

Since the ultimate goal in protein folding is to understand how protein folding takes place *in vivo*, the consideration of celluar chaperones is of great importance. However, several further relevant aspects have to be considered for a deeper understanding of cellular protein folding [22]. Besides the fact that in the cell proteins typically start to fold already during their synthesis (co-translational folding) the folding in crowded conditions is accompanied by many possible interactions, including interplays with the ribosome and ribosome-associated proteins. In this respect the coupling between the synthesis rate of aminoacids of the polypeptide chain and the folding of the nascent chain is most probable the most significant difference to the classical *in vitro* refolding scenario [23].

References

[1] J. Buchner and T. Kiefhaber, *Protein Folding Handbook Vol. 1-5*, (Eds.: J. Buchner, T. Kiefhaber), Wiley VCH-Verlag, Weinheim, 2005

[2] B. Nölting, *Protein Folding Kinetics*, Springer, Berlin, 1999

[3] K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikel, *Annu. Rev. Biophys.* **37**, 289-316 (2008)

[4] S.E. Radford, *Trends Biochem. Sci.* **25**, 611-618 (2000)

[5] J. Kuriyan, B. Konforti, and D. Wemmer, *The molecules of life: physical and chemical principles*, Garland Sciences, Taylor & Francis Group, New York, 2013

[6] E. Braselmann, J.L. Chaney, and P. L. Clark, *Trends in Biochem. Sci.* **38**,337-344, 2013

[7] H. Eyring, *J. Chem. Phys.* **3**,107-115, 1935

[8] J. Kubelka, J. Hofrichter and W.A. Eaton, *Curr. Opin. Struct. Biol.* **14**,76-88, 2004

[9] H.S. Chung, K. McHale, J.M. Louis, W.A. Eaton, *Science* **335**,981-984, 2012

[10] S.W. Englander and L. Mayne, *PNAS* **111**,15873-15880, 2014

[11] J. Stigler and M. Rief, *Science* **334**,512-516, 2011

[12] M.Pirchi, G, Ziv, I. Riven, S.S. Cohen, N. Zohar, Y. Barak, and G. Haran, *Nat. Comm.* DOI:10.1038/ncomms1504, 2011

[13] V.N. Uversky, *Protein Science* **22**, 693-724, 2013

[14] S.J. Demarest et al., *Nature* **415**, 549-553, 2002

[15] T. Kiefhaber, A. Bachmann, and K.S. Jensen, *Curr. Opin. Struct. Biol.* **22**, 21-29, 2012

[16] T. Lee, C.R. Moran-Gutierrez, A.A. Deniz, *Semin. Cell Dev. Biol.* DOI:10.1016/j.semcdb.2014.09.027 ,2014

[17] F.U. Hartl and M.Hayer-Hartl, *Nature Struct. Mol. Biol.* **16**, 574-581, 2009

[18] A. Mashaghi, G. Kramer, D.C. Lamb, M.P. Mayer, and S. J. Tans, *Chem. Rev.* **114**, 660-676, 2014

[19] F. Gloge, A.H. Becker, G. Kramer and B. Bukau, *Curr. Opin. Struct. Biol.* **24**, 24-33, 2014

[20] S. Sharma, K. Chakraborty, B.K. Müller, N. Astola, Y. Tang, D.C. Lamb, M. Hayer-Hartl and F.U. Hartl, *Cell* **133**, 142-153, 2008

[21] A. Mashaghi, G. Kramer, P. Bechtluft, B. Zachmann-Brand, A.J.M. Driessen, B. Bukau& S.J. Tans, *Nature* **500**, 89-102, 2013

[22] K.S. Hingorani and L.M. Gierasch, *Curr. Opin. Struct. Biol.* **24**, 81-90, 2014

[23] A. Katranidis, D. Atta, R. Schlesinger, K.H. Nierhaus, T. Choli-Papadopoulou, I. Gregor, M. Gerrits, G. Büldt, J. Fitter, *Angew. Chem.* **48**, 1758-1761, 2009

E 3 Protein Dynamics

R. Biehl
Institute of Complex Systems, JCNS-1 & ICS-1
Forschungszentrum Jülich GmbH

Contents

1. Introduction 2

2. Local movements 3

2.1. Atomic vibration 3

2.2. Sidechain movements with Atomic Resolution: Time Resolved X-ray Crystallography. 4

2.3. Microsecond dynamics of Sidechains and Picosecond Dynamics of the Backbone observed by Nuclear Magnetic Resonance Spectroscopy 5

3. Domain motions 10

3.1. Phosphoglycerate kinase as a classical hinge 11

3.2. Intrinsically Disordered Proteins (IDP) 12

3.3. Cooperative Rotation of the F1-ATPase Motor 14

4. Summary 15

References 16

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

Proteins are biological macromolecules present in all cells and in body liquids. They work as nanomachines of live to produce material, move objects to the place where they are needed, degrade toxic chemicals, regulate the velocity of processes or protect the cell e.g. from viruses as part of the immune system. Proteins are synthesized as linear polypeptides of 21 different amino acids that are connected by peptide bonds. The primary structure of the protein as the sequence of amino acids is coded in the sequence of DNA. The side chain defines the amino acid properties as acidic or base, hydrophilic or hydrophobic, polar or nonpolar. During the protein synthesis the protein strand folds to a unique 3 dimensional structure (e.g. α-helices, β-sheets, disordered regions, hairpin). The structures is mainly stabilized by hydrogen bonds, hydrophobic amino acids in the core of a domain and hydrophilic amino acids at the surface.
Protein domains are preserved regions of the protein structure with hydrophobic core and hydrophilic shell and can evolve function even if separated from the protein. The domain size reaches from 40 amino acids to several hundred amino acids with an average of approximately 100 [1,2]. Domains of similar structure are found in different species and similar functions show often similar domain structure, which is conserved during evolution and one speaks of domain families conserved during molecular evolution [3].
Activity of proteins is often related to an active center where actual function takes place. This can be catalysis of a chemical reaction, binding of a substrate or the change of a chemical potential. An early model about protein activity and specificity was the “lock and key” model, which assumes an exact fit of the protein active site to the substrate due to complementary geometrical shapes but with a rigid conformation as found in crystal structures [4]. To explain also the stabilization of the transition state with bound substrate in a different configuration compared to the unbound state the later “induced fit” model [5] allows a reshaping of the binding site to the substrate including local configurational changes of amino acids or large structural changes as for allosteric transitions. Still the protein is viewed as a rigid structure in liganded and unliganded case. Today it is realized that proteins are quite flexible objects, which show configurational changes on all length- and timescales. To reach a buried active site it is often necessary to open a cleft that the substrate can enter and to release the product. Binding of substrates in specific pockets allows to bring them close together in a specific configuration for e.g. phosphate or hydrogen transfer in a protected environment. Necessary conformational changes can be the rate-limiting step in catalysis. In other cases as for kinesin walk on myesin the configurational change is the aim of a chemical process to allow transport of cargo [6]. Therefore protein dynamics on all length scales is a key to understand how conformational changes are related to function and which mechanisms are involved to allow the rich functionality of proteins.
There are two different but linked types of dynamic motions. In terms of an energy landscape view thermal motions are motions that cover the configurational space at thermal energy kT in equilibrium [7]. The accessible configurational space can spread over a single deep minimum or a broader rugged valley where in both the depth defines the occupancy of a configuration in the valley within Boltzmann distribution. Kinetic motions try to find the equilibrium from a higher energy level and are directed towards equilibrium. The higher energy level can be due to an excitation (e.g. photolysis) or binding of a ligand that changes the local energy landscape. Nevertheless thermal motions occur also in the excited state and may help to overcome energy barriers on the kinetic pathway.
The fastest motions in proteins are bond vibrations, side chain rotation at the protein surface or torsion of buried methyl groups with sub-angstrom amplitudes on picosecond timescales. Rearrangements of amino acids to adjust the orientation of functional groups may require local flexibility of neighboring amino acids in a cooperative manner that slows down the

process. Movements of secondary structure elements or rearrangement of groups of amino acids on nanosecond timescale allow the adaption of the protein structure to bind specific ligands. Slower motions with larger angstrom amplitudes are relative motions of complete domains as hinge bending movements or swapping of domains depends strongly on the local environment and can be fast as several nanoseconds or slow as up to seconds dependent on the needed rearrangements and the involved interactions. Allosteric transitions, functional conformational changes, folding and unfolding will happen on microsecond timescale and nanometer length scales. In general all these motions are dependent on each other and are coupled. The local atomic fluctuations lubricate the domain motions on larger length scales and domain motions change the shape of the protein.
In the following an overview over some configurational movements is given together with a basic explanation of experimental techniques allowing detecting timescale and amplitude.

2 Local movements

Local movements comprise movements of single atoms or small atom groups. Due to thermal excitation bonded atoms show vibrating movements as eigenmodes of their specific configuration. Each sidechain of an amino acid can have different configurations with respect to the backbone dependent on the space required for a change e.g. in orientation. Amino acids at the surface have more configurational freedom compared to completely buried amino acids. Configurational changes can be due to thermal movements or due to specific processes as binding of a substrate.

2.1 Atomic vibration

The fastest movements with highest energy are atomic vibrations as they are common for any molecule. Fig.1a shows as an example the geometry of torsion around the C-C bond of a methyl group, a symmetric stretching of the C-H bonds and a symmetric bending of the c-H bonds of the sidechain of alanine. These are only a few possible vibrations of alanine sidechain and each amino acid has different vibrational frequencies dependent on the atomic structure. For an overview see Barth et al [8]. The exact frequency depends not only on the geometry as for free molecules of the same architecture, but also on the direct environment and neighboring amino acids. Hydrogen bonds or polar interactions alter the vibration frequency. Fig.1c shows an example spectrum of $(PPG)_n$ an synthetic polypeptide with a similar structure to natural collagen [9]. The frequency range reaches from 100 cm^{-1} to about 2000 cm^{-1} (12.4 meV – 250 meV) describing motions on a timescale of 0.01ps to 0.3 ps. The difference in the spectra is related to a partial deuteration that changes the frequency and the amplitude of specific vibrations due to the change of the hydrogen mass. In this way the specific exchange allows to separate some of the vibrational modes present in the sample. Typical instruments with the necessary large energy transfer are neutron time of flight instruments.
For proteins the Amide I absorption band between 1600 cm^{-1} and 1700 cm^{-1} is of importance because it allows the determination of the relative content of secondary structure in the protein. It can be measured by infrared spectroscopy (e.g. FTIR). In this absorption band the C=O stretch vibration of the peptide backbone is dominant [10]. As shown in Fig.1b the hydrogen bond between the C=O and the N-H of a different amino acid stabilize the secondary structure elements. A specific hydrogen bond modifies the stretching vibration in a way that is characteristic for the local geometry defined by the secondary structure.

Measuring the different contributions allows extracting the fractional content of secondary structure elements.

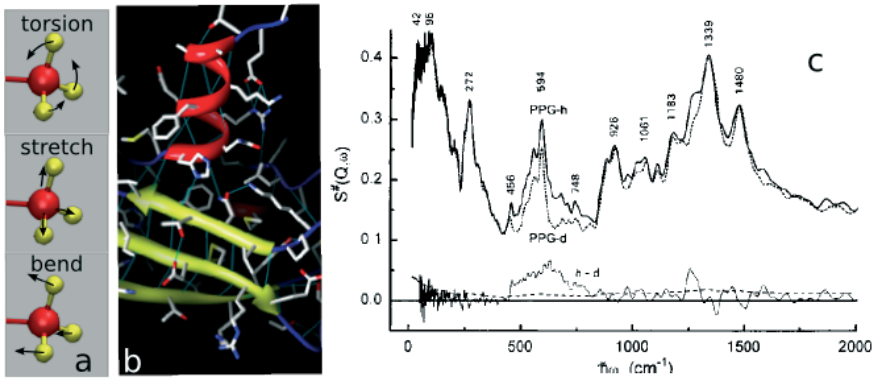


Fig. 1: *Example geometries of torsion, symmetric stretching or bending vibrations of a methyl group as the side group of alanine. b) Secondary structure elements* β*-sheets and* α*-helices (in yellow, red) with side chain elements. The secondary structure is stabilized by the hydrogen bonds (thin blue lines). c) Vibrational spectrum of synthetic polypeptide (PPG)n associated in a right-handed supercoiled triple helical arrangement as found in natural collagen measured at the time-focused crystal analyzer spectrometer (TFXA) at ISIS, UK by Middendorf et al. [9].*

2.2 Sidechain movements with Atomic Resolution: Time Resolved X-ray Crystallography.

Conventional x-ray crystallography measures the diffraction patterns in different orientations of a single crystal with hundreds to thousands of Bragg- reflexes from which the three-dimensional electron density is calculated. The crystal structure of atom positions is generated as found e. g in the Brookhaven Protein Databank (PDB). Configurational changes due to activity can only be examined if it is possible to grow a crystal in both states as e.g. liganded with the substrate or an inactive replacement of the substrate and without the substrate. Diffusion trapping can be used to find the binding site of a substrate that is able to diffuse into a substrate free crystal. A larger configurational change due to substrate binding is difficult to access in this way and the timescale of a process cannot be accessed.

Time resolved crystallography tries to measure diffraction patterns with a delay time to observe the time evolution of a process. To access a dynamical process the protein must be active in the crystalline state and the process needs to be triggered inside the crystal with some reasonable amount of concentration. One way is to use a pump-probe method with measurements after a trigger event. Measurements with defined time delay relative to the trigger event e.g. photolysis of a process by a laser pulse allows to follow the process with different delays. On the other hand a complete series with defined delays can be acquired if the repetition rate for a measurement is high enough. To overcome the need of different orientations for sub-second resolution a polychromatic Laue X-ray diffraction technique can be used [11]. Larger motions as domain movements cannot be observed as the crystal structure limits the configurational freedom to move over larger distances.

Schotte et al. have reported 100 picosecond time resolution x-ray crystallography of a myoglobin mutant after photolysis by an orange laser flash [12]. Fig.2A shows the electron density map of the unphotolyzed myoglobin and 100ps after the flash showing 3 larger configurational changes (arrows) and distributed smaller changes in the protein.

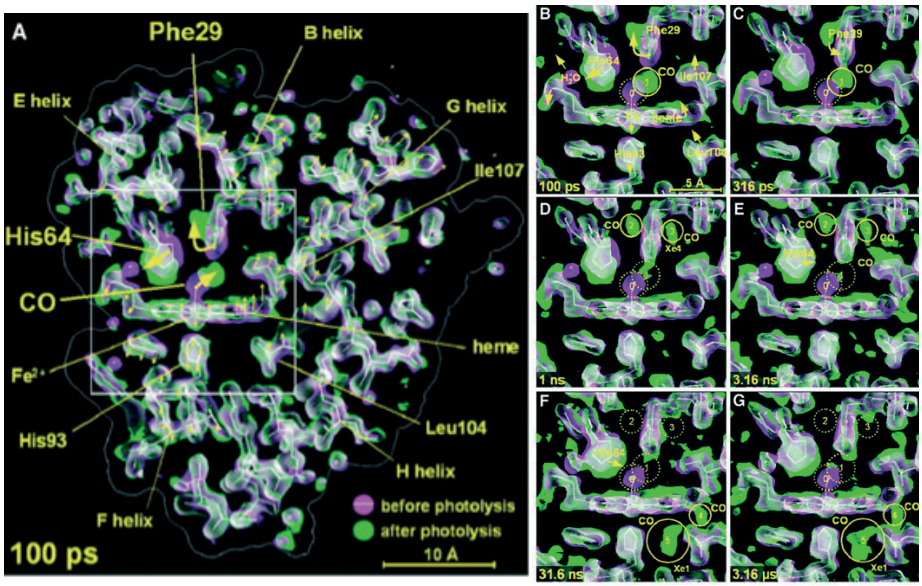


Fig. 2: *Electron density map of myoglobin before (magenta) and after photolysis (green). Overlapping densities are shown in white. The stick model included shows the unphotolyzed state. Large arrows indicate 3 large changes while small arrows indicate small rearrangements in the whole protein. Sequence of an enlarged view of A. B-E) Times are 100 ps, 316 ps, 1 ns, 3,16 ns, after the laser flash. F and G are 31.6ns and 3.16 µs after a longer ns laser flash. Circles indicate location of CO molecules. Figure from [12].*

Fig.2 B-E shows a sequence of delay times after the laser flash. The CO molecule dissociates and is trapped 2Å apart from the original binding site (positions 2 + 3). Phe29 is displaced but relaxes back to its original position within 316 ps. After several nanoseconds His64 relax to their deoxy position and the CO molecule has migrated to position 4 and 5 where it is trapped for several microseconds.
Here the conformational change due to photolysis is demonstrated with a correlated motion of sidechains rapidly sweeping away the CO molecule from its preferred binding site.
The experiment demonstrates that fast sub nanosecond reorientations are possible even in a restricted Crystal structure.

2.3 Microsecond dynamics of Sidechains and Picosecond Dynamics of the Backbone observed by Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance measures the resonance frequency characteristic for a spin flip in a magnetic field, which depends on the respective atom type (^{1}H, ^{13}C, ^{15}N) and e.g. the local electronic environment leading to the chemical shift (dependent on binding partners, bond length and angles) or J-coupling (interaction of different spins). Chemical exchange

phenomena as conformational changes, exchange with the solvent or ligand binding modifies this local environment. Fig. 3 shows as an example for a two site chemical exchange the chemical shift dependence on the rate of exchange relative to the measurement time.

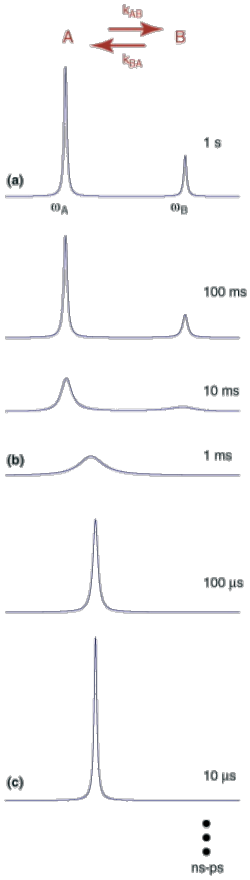


Fig. 3: *An overview over timescale in NMR for the case of a two site chemical exchange with a population of 3:1 between A and B. The difference in chemical shifts* $\boldsymbol{\Delta\omega \approx \omega_A - \omega_B}$ *=100Hz determines a "shutter time" as the NMR time scale by* $\boldsymbol{\sqrt{2}/\pi\Delta\omega}$*. a) The exchange time* $(k_{AB}+k_{BA})^{-1}$ *(indicated as times in the figure) is slower allowing to detect 2 separate peaks with intensities related to population of A and B. b) The exchange is on same timescale showing coalescence to a single broadened peak at population averaged position, because within the measurement time single transitions are likely. c) The exchange is much faster and a single sharp peak is observed. Within the measurement time the transition occurs several times. The position depends on population of A and B. For known chemical shifts of A and B the peak position can be used to determine the populations. Figure used from [13].*

Microsecond-millisecond changes can be detected due to modification of the chemical environment and lead to a change of the specific chemical shift. Molecular motions on picoseconds to nanoseconds as sidechain vibrations can be detected by measuring spin relaxation.
2D-NMR measures the coupling between spins of different atoms in a molecule. Methods like correlation spectroscopy (COSY), ECOSY, HSQC, EXSY allow the detection of correlations between atoms that are connected over several bonds. E.g. Nuclear Overhauser effect spectroscopy (NOESY) allows detecting correlations that are not connected by a bond, but which are not to far separate in space.

Fig. 4 shows as an example magnetization exchange spectroscopy (EXSY) for the observation of a slow process on the second timescale. For different configurations two peaks are observed with a cross peak as a direct indication of transition between configurations within the measurement time.

Measuring with a delay time allows the determination of rate constants related to the transition. Faster motions on the pico-nanosecond timescale are accessible by spin relaxation spectroscopy. The relaxation rate depends on molecular motions of atoms in the local environment that interact via dipole-dipole interactions or by chemical shift anisotropy and the global rotational motion [13]. Bond vector motions (e.g. amide N-H bond of the backbone in a ^{1}H-^{15}N experiment) are separated into contribution from global rotation on the timescale of nanoseconds, which is equal for all residues, and internal dynamics on the picosecond

timescale to extract local backbone dynamics. Limitations arise when the dynamics of internal motion becomes slow and reaches the timescale of rotational motion.

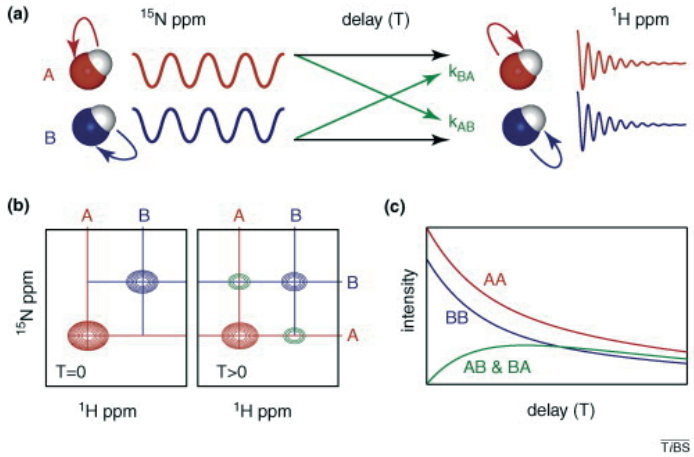


Fig. 4: *Magnetization exchange spectroscopy (EXSY) can observe the slow exchange between conformational states A and B. In a standard correlation experiment observing the backbone amide, the magnetization is transferred from ^{1}H to ^{15}N, the ^{15}N chemical shift is measured and afterwards the magnetization is transferred back to measure the ^{1}H magnetization by complex RF-pulse sequences. Dependent on the configuration correlation peaks A or B are observed (equivalent to b) T=0). In the EXSY experiment the ^{1}H experiment is done after a delay time T>0 (see a) and a cross peak is observed in cases were during the delay time a transition between A and B occurred (b T>0). Varying the delay time allows measurement of the relaxation between the two configurations A and B. c) The time dependence of a relaxation with decreasing intensities for A and B. The cross peak intensity increases at short times. All intensities decrease at long times because of spin relaxation. From the relaxation the forward and reverse rate constants can be extracted. The technique is applicable for rate constants in the range from 0.5 s^{-1} to 50 s^{-1}. Figure used from [13].*

Configurational Transition in ClpP on Second Timescale

A slow transition in the oligomeric protease ClpP, consisting of 14 subunits arranged in a heptameric ring with a total mass of 300kDa, was observed by magnetization exchange spectroscopy (EXSY)[15]. ClpP is a cylindrical self-subdividing protease with a chamber containing the proteolytic active site were through axial pores the already unfolded substrate is inserted and proteolysis takes place (see Fig.6a).

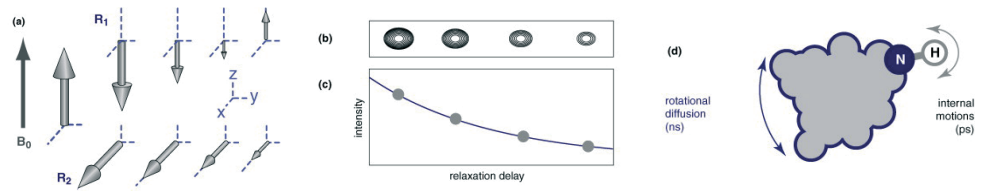


Fig. 5: *Spin relaxation: a) At equilibrium the spins are aligned to the magnetic field. After a specific perturbation (RF pulse) the spin is aligned antiparallel for longitudinal magnetization R1 or perpendicular for transverse magnetization R2. After a delay time allowing relaxation an inversion pulse is applied and after a further delay the non-relaxed magnetization can be measured. b+c) The magnetization is interpreted in a model free approach by* $S^2+(1-S^2)\exp(-t/\tau)$ *[36]. c) R1 relaxation depends on global tumbling e.g. due to rotational diffusion and local fluctuations. R2 depends only on the local relaxations. Local relaxations can be extracted by combined measurements. Figure used from [13].*

The assignment of the δ1 methyl peaks for I149 and I151 shown in Fig. 6b was done by site directed mutagenesis and presents two peaks for each of the residues, a broad and a narrow peak indicated as F and S for fast and slow relaxing. By EXSY it was established that the two peaks correspond to two different configurations and that a transition occurs with a rate constant of 60 s^{-1} at 0.5 °C. Both residues show the same rate constants suggesting that the same process is observed. I149 and I151 are present in all monomers building a ring at the connecting interface between the heptameric rings. The observed configurational change could be related to opening of pores, which allow the release of the product after proteolysis. This was tested by mutagenesis in introducing at position 153 a cysteine that can build a disulfide bridge between monomers if oxidized. A series of kinetic measurements showed fast product release from the reduced form, but product release was not observed in the oxidized form. The disulfide bridges quench the release process. The relaxation process observed by NMR is likely related to the release of the proteolysis products through pores in the equatorial plane.

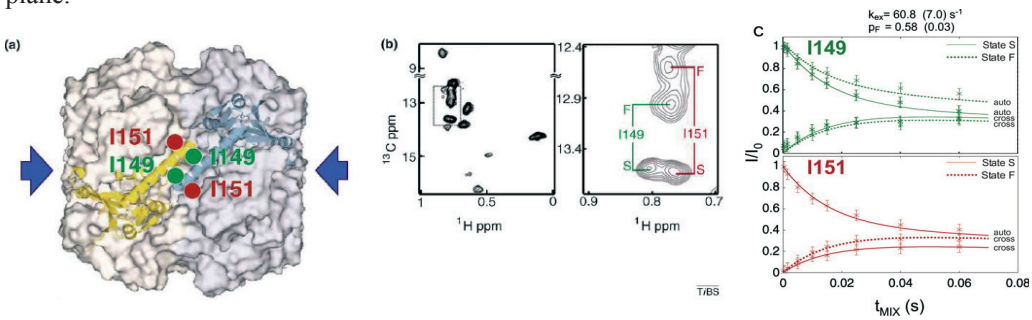


Fig. 6: *a) ClpP Protease with two monomers shown in yellow and blue. Isoleucines are indicated as red and green circles. Blue arrows indicate substrate-entering pores. b) TROSY correlation spectrum with the two δ1 methyl peaks (indicated F and S for fast and slow in the magnification on left of the small rectangle right) of I149 and I151 residue. The occurrence of two peaks is a result of the slow exchange between two conformations. c) Relaxation of the peak intensities with the relaxation of the cross peak. All relaxation rates are approximately equal. Figures from [14,15].*

By EXSY it was established that the two peaks correspond to two different configurations and that a transition occurs with a rate constant of 60 s^{-1} at 0.5 °C. Both residues show the same rate constants suggesting that the same process is observed. I149 and I151 are present in all monomers building a ring at the connecting interface between the heptameric rings. The observed configurational change could be related to opening of pores, which allow the release of the product after proteolysis. This was tested by mutagenesis in introducing at position 153 a cysteine that can build a disulfide bridge between monomers if oxidized. A series of kinetic measurements showed fast product release from the reduced form, but product release was not observed in the oxidized form. The disulfide bridges quench the release process. The relaxation process observed by NMR is likely related to the release of the proteolysis products through pores in the equatorial plane.

Amide ^{15}N Backbone Dynamics in Adenylate Kinase

Atomic fluctuations on picosecond-nanosecond timescale can facilitate larger motions on slower timescale as was shown for adenylate kinase by spin relaxation spectroscopy of the ^{15}N bonds by Henzler-Wildman et al. [16]. A mesophilic and thermophilic homologue were examined at different temperatures. The measured spin relaxation spectra were analyzed by models derived from model free analysis, determining the order parameter S^2 and a relaxation time from the relaxation curve of each residue. In the basic model relaxation is described by $C(t)=S^2+(1-S^2)\exp(-t/\tau)$. The S^2 describes the change from completely rigid structures ($S^2=1$) over flexible structures (0.85 for secondary structure or 0.5 for unstructured regions) to completely uncorrelated motions ($S^2=0$) and is most sensitive to local packing [17]. Different models and additional measurements were used to remove contributions from lid movements on µs scale, anisotropic diffusion or coupled motions. Rotational correlation time was found as 14 ns^{-1} and 18 ns^{-1} at 20°C for thermophilic adenylate kinase in open and closed configuration in accordance to calculation of HYDRONMR [18] based on the PDB structure.

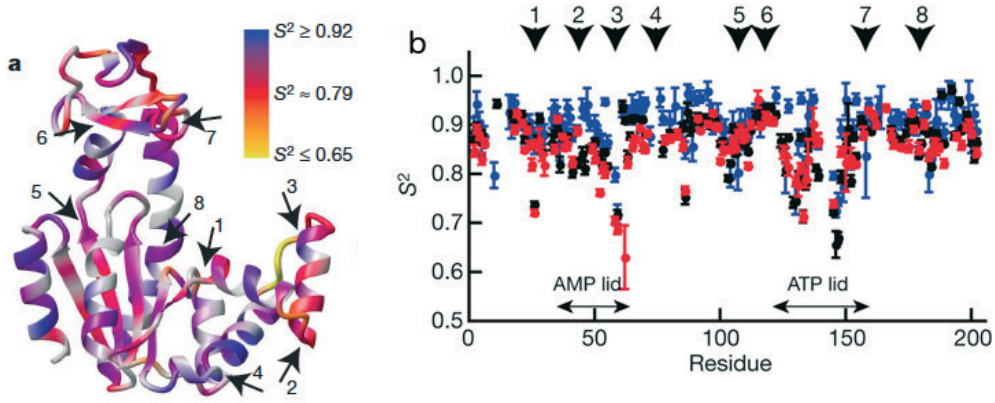


Fig. 7: *a) Flexibility of the mesophilic adenylate kinase at 20°C. Colors indicate the value of the order parameter S^2 with grey for residues for which S^2 cannot be measured (due to fast hydrogen exchange or spectral overlap). b) Temperature dependence of S^2 for the thermophilic adenylate kinase at 20°C (blue), 50°C (black) and 80°C (red). Numbers with arrows indicate in both plots identified hinges. Figures from [16].*

Fig.7a shows the mesophilic adenylate kinase colored according to the value of S^2 with values between 0.92 as rigid and 0.65 for flexible regions. It was found that hotspots of flexibility are found where backbone conformation must change for lid motion as indicated by arrows in the figure. E. g. hinge 8 corresponds to a kink in the long α-helix if the lid is closed during substrate binding. Fig. 7b shows the S^2 values for all residues with marked hinges. With increased temperature the overall flexibility increases. At 20°C the thermophilic adenylate kinase shows smaller order parameter values compared to the mesophilic homologue. Similar order parameters for both homologues are found if both are measure in the same distance from their living temperature (30° below 54°C and 109°C), supporting the hypothesis that fluctuations are related to stability.

3 Domain motions

Domain motions are correlated motions between different domains or inside of domains like bending and torsion of the domain. The mechanisms are described as shear motions along an interface or hinge motions where such an interface is missing, but unclassifiable motions are allowed. The connection between domains can be a broad soft hinge as in the case of phosphoglycerate kinase, a single α-helix as in the case of lactoferrin or a disordered amino acid sequence of the protein chain as in the case of immunoglobulin G1 or mercury ion reductase.

Configurational changes changing the shape of a protein by rearranging complete domains can be observed by methods, which allow the detection of correlated motions over larger distances. Förster resonance energy transfer (FRET) measures the distance distribution of chromophores attached to cysteines at specific sites on the protein surface or introduced by site directed mutagenesis [19]. The excitation of the donor is followed by a transfer of the energy to the acceptor – overlap of absorption and emission spectra is essential - and detection of the later emission. The efficiency of energy transfer is dependent on the distance and allows measuring a distance distribution between the chromophores respectively the points where the chromophores are bound to the protein. Single molecule fluorescence polarization anisotropy (smFPA) measures dynamic changes of the orientation via time resolved polarization measurements and yields information about size, shape and rotational dynamics [20]. Electron spin resonance can be used to measure the interaction between two unpaired spins similar to NMR, but because of the stronger interaction the distances can be larger [21]. Because of the absence of unpaired electrons in proteins, spin labels have to be inserted by site directed mutagenesis. Another technique is small angle scattering by neutrons or x-rays (SANS, SAXS). Both techniques allow examination of the low-resolution structure and the ability to compare structural changes in solution due to changes of pH, salt concentrations, temperature or substrate addition. Time resolved SAXS can reach sub-millisecond resolution and can be combined with stopped flow experiments or trigger events as in time resolved x-ray crystallography [22].

Neutron spin echo spectroscopy (NSE) is a technique that is able to access the timescales from 0.1 up to several hundred nanoseconds and simultaneously covers the length scale relevant for protein domain movements as in SAXS or SANS of several nanometers distance between domains [23]. NSE measures the temporal correlation of configurational changes under utilization of the neutron spin to detect tiny velocity changes during the scattering process. The measured intermediate scattering function can be interpreted as a time correlation between small angle scattering patterns with nanosecond resolution. Main

contributions to the correlation come from translational and rotational diffusion due to the spatial correlation of different proteins diffusing in the solution and internal domain dynamics on nanosecond timescale. In the following we present exemplary results to demonstrate a small variety of possible motional patterns.

3.1 Phosphoglycerate kinase as a classical hinge

Phosphoglycerate kinase (PGK) is an enzyme that is involved in glycolysis. It relocates a phosphate group from 1,3-biphosphoglycerate, an intermediate product in glycolysis, to ADP to synthesize ATP [24]. PGK is composed of two separated domains connected by a hinge region as shown in Figure 8 [25]. 1,3-biphosphoglycerate and ADP are bound at opposite positions in the cleft at the two domains. The active site is located at the hinge between the C-terminal and N-terminal domains. Evidence for a hinge bending motion induced by substrate binding was found by Bernstein at al. by comparing crystallographic structures of different species with and without bound substrates bringing the substrates closer together [26]. The crystal structure without substrate has an open cleft configuration with key residues Arg-39 and Gly-376 in the active center separated by about 1.18 nm. The proposed mechanism of induced fit due to substrate binding closes the cleft by a 32° hinge closure to the active configuration with bound substrate. This cleft closing motion mainly brings key residues Arg-39 and Gly-376 of the active center together with the substrates to a distance of 0.35 nm as found in the closed cleft crystal structure.

Fig.8 shows SANS measurements of PGK in solution with and without substrate demonstrating that the solution structures are different from the substrate bound crystal structure [27]. Modeling the structure by deformations along softest elastic normal modes (torsion and 2 perpendicular bends of the hinge) allowed modeling of the deformation due to substrate binding in solution. The distance between the active residues was reduced from 1.14 nm without substrate to 0.82 nm with bound substrate, but still to far to allow activity within a static structure.

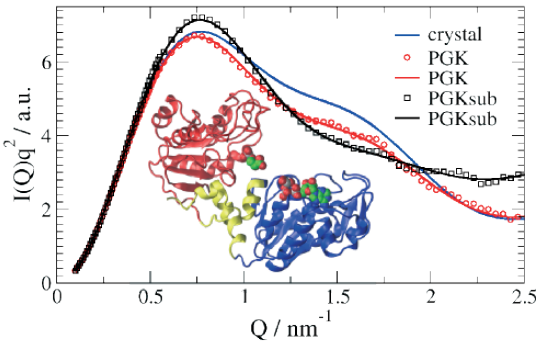


Fig. 8: *Form factor measured by SANS from PGK and PGKsub (Kratky plot: Q versus $Q^2I(Q)$). Lines show the calculated form factor from the crystal structure (blue) and from structures deformed along the softest normal modes to fit the experimental data. The inset shows the protein with the hinge in yellow, the main domains in blue and red and the substrates as spheres. Figure from [27].*

NSE measurements show the relaxation of the intermediate scattering function dependent on diffusion and internal dynamics as shown in Fig.9 left. At low scattering vectors Q (observing large length scales) the protein looks point like and a single exponential relaxation is

observed. t larger Q (observing length scale of protein size) additional contribution arise from internal dynamics. At long times diffusion is observed, that can be described within a single model for all Q values by rotational and translational diffusion constants. The additional component at short times is described by a Q dependent amplitude and a relaxation time τ, which both depend on the construction of the specific hinge and influence function.
The observed relaxation times are 60 ns for the substrate free and 45 ns for the substrate bound PGK. Fig.9 shows at the right the amplitude A(Q) with model calculations based on the softest normal mode deformations. From the deformation of the hinge the distance of the active residues can be calculated. It was found that the thermal driven deformation of the protein at the hinge is strong enough to reach configurations that allow activity as the active residue come close enough together.
In summary, the NSE investigation has demonstrated that the approach to a functional configuration of PGK needs to be attributed to the dynamic fluctuations of the main domains instead of an earlier proposed induced fit by substrate binding. Thus, in the case of PGK, hinge dynamics enables function.

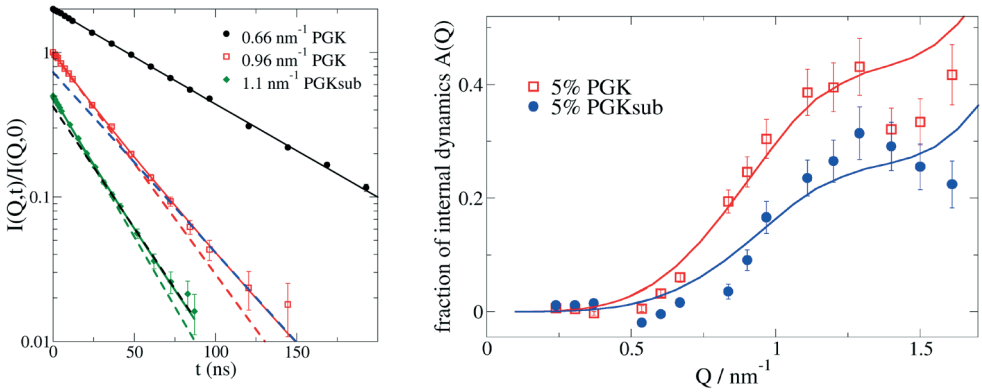


Fig. 9: *Left: Semi logarithmic plot of I(Q,t)/I(Q,t=0) for selected Q values (PGK, black and red; PGKsub, green; data are shifted for clarity and are equal 1 for t=0). Red and green dashed lines represent the initial slope extrapolated to long times. The blue and black dashed lines represent the long-time limit extrapolated to t=0. The long-time limit corresponds to rigid-body diffusion including inter-particle effects. The difference between extrapolated long time diffusion at t=0 to the initial slope amplitude is the internal dynamics contribution A(Q)exp(-t/τ). Right: Q dependence of A(Q) compared to model calculations Figure from [27].*

3.2 Intrinsically Disordered Proteins (IDP)

About 40 % of all proteins in the human body are intrinsically disordered. These IDP's do not exhibit any well-defined folded structure that could be crystallized. Their structural and dynamic properties reach from very soft structures over folded elements connected by extended and flexible loops to fully disordered polypeptide chains. The biological role of the IDP is founded in their high conformational adaptivity, enabling them to respond rapidly to environmental changes or other macromolecules allowing them to fold into different states or even a rigid 3D structure. For these properties dynamics is essential. Specifically the sampling of the energy landscape and the exploration of the large conformational space are driven by

conformational motions of the unfolded peptide chain. On the other hand IDPs show the same dynamics as proteins during the early stage of folding. Since the intrinsic disorder prevents crystallographic structure determination, only low resolution SANS and SAXS information about the average structure in the disordered state exists.
Myelin basic protein (MBP) is a major component of the Myelin sheaths in the central nervous system [28]. In the human body MBP is of significant importance as there are many neurological disorders, as e.g. multiple sclerosis, that are related to MBP mal function. Lipid free MBP is not completely unfolded but retains some elements of the alpha helix and beta sheet (about 60 % of the protein is unfolded) [29].

Fig.10a+b display X-ray form factors of MBP that display strong similarities to polymer form factors as the high Q data show a power law close to Q^{-2} that is characteristic for Gaussian chain polymers in theta solvents. The small increase visible in the Kratky plot at high Q indicates a length scale where the random oriented character vanishes and the linear character of short chain segments becomes visible. A Monte Carlo simulation was used to generate a coarse grained ensemble representing the structural characteristics of and a ensemble was selected that represent the SAXS data [30]. The resulting characteristic molecular shapes are displayed in Fig.10c. The model conformations indicate an elongated structure with a relatively compact core and flexible ends on both sites. Fig.11 at left displays NSE spectra from 54 mg/ml solutions (times up to 140 ns could be accessed). Inspecting this figure the two component structure of the NSE spectra at Q-values above 0.9 nm^{-1} is visible. Thus, we deal with long time rigid body motion augmented by internal dynamics with relaxation times below 10 ns.

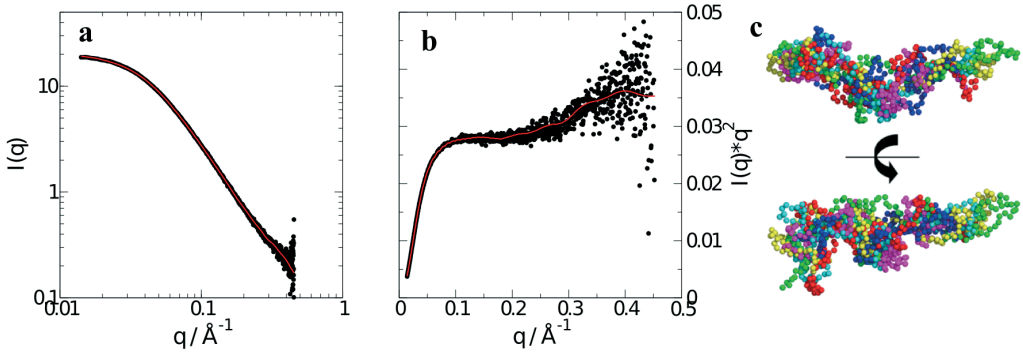


Fig. 10: *SAXS data from MBP at 4.5 mg/ml. a) The red solid lines is a fit with the Debye equation for a Gaussian chain. (b) Kratky plot. The line is a result of the scattering from the most probable conformational ensembles shown in c. c) Representative coarse-grained conformations of MBP as determined by inverse Monte Carlo. The structures are rotated by 90 °C in the lower part of the figure. The color code relates to six different realizations of the ensemble, which represent the data best. Figure from [31].*

The structural models based on SAXS analysis were used to describe the long time translational and rotational diffusion combined with a Q-dependent motional amplitude A(Q) and an internal mode relaxation time. The characteristic internal relaxation time τ_{int}=8.4±2.0 ns is found for the whole structural ensemble and the corresponding amplitude A(Q) is

displayed in Fig.11 at the right. Normal mode analysis was used to describe the deformation of the structural models. Fig.11 middle shows the first two normal modes as a bending of the structure, which already give a satisfactory description of the observed amplitudes. Comparing the normal modes with the structural models in Fig. 10c it can be concluded that the normal modes describe approximately the motion from one structural model to the next. Here it is shown that even for very flexible structure as the amino acid chain still low frequency collective stretching and bending motions of the outer part of the structure describe the essential features of the large-scale dynamics.

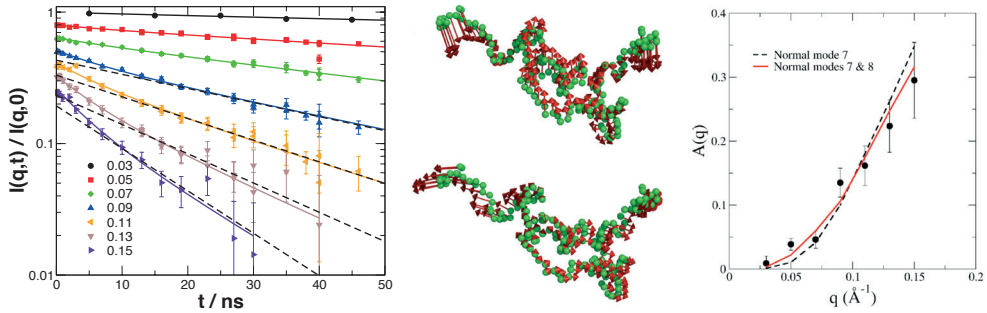


Fig. 11: *Left) NSE Data for MBP: All spectra start at unity but are shifted consecutively by a factor of 0.8 for clarity shown up to 50 ns. Solid lines are fits to the NSE data with the structural model. The dashed lines are exponential fits for t>20 ns to extrapolate the long time rigid body dynamics. A clear separation between the internal and the global dynamics is obvious. Middle) Displacement pattern of the normal modes 7 (upper part, bending) and 8 (lower part, stretching) from the structural model according to normal mode analysis. The lengths of the vectors are increased for better visibility. Right) Amplitude of the internal protein dynamics as obtained from the fit. The solid and dashed lines are the calculated mode amplitude according to equation 4. Figures from [31].*

3.3 Cooperative Rotation of the F1-ATPase Motor

Adachi et al. demonstrated the stepping rotation of the F-ATPase motor through angle-resolved smFPA [32]. F-ATPase is a membrane protein complex that synthesizes ATP. A proton gradient between membrane sides is used to drive the reaction as the protons cross the membrane in a transport reaction in the membrane bound F_0-ATPase, which acts as a rotor. F_0-ATPase reaches into the stator F_1 by a shaft (dotted line, see Fig.12). A fluorescent probe was attached on top of F_0 reaching through F_1 to monitor the polarization of emitted or absorbed light during activity.

Fig.12 shows on the right the intensity course during activity with the calculated orientation and corresponding revolution angle. The time evolution of a continuous motion is shown for comparison as a red line.

The stepwise character of the motion is demonstrated as a true property of the ATPase and further analysis results in 120° steps with a dwell time dependent on ATP concentration in the solution as e.g. 5.5s at 20nM ATP.

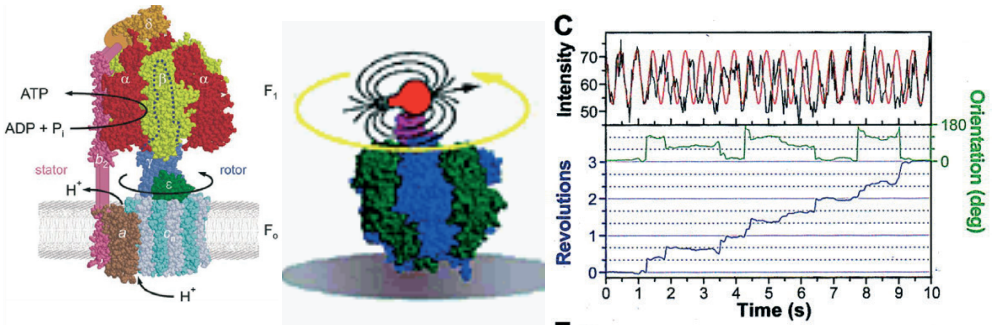


Fig. 12: *Left) Schematic sketch oft the F-ATPase complex in the membrane with F_1 in yellow-red and F_0 in light-blue, green in the membrane. The rotor (blue) goes through F1 up to the top. (http://uni-marburg.de/uT9Bj). Middle) F_1-ATPase motor with a single fluorescent probe attached to the rotor reaching through F_1. Right) Time trajectories of the fluorescence intensity (black, top) and calculated fluorophore angle between 0° and 180° (green, bottom). The accumulated rotation angle (blue, bottom) was obtained by assuming that all steps were counterclockwise, reprinted from [20,32].*

4 Summary

In the past the function of biological assemblies was discussed in terms of structure, which was in most cases derived by X-ray crystallography. In recent years the biological community became more and more aware of the importance of motions and dynamics in proteins that can play an important role in understanding function.

The importance of protein dynamics may be highlighted in the frame of drug design. While in the past in general the development of drugs was done using static crystallographic structures implying the lock-and-key model, during the last 10 years the state-of-the-art involves ensemble docking. Ensemble docking considers conformational changes and searches for conformations where a drug can bind to the active site of the protein, meaning that metastable protein states identified in molecular dynamics (MD) simulation are individually targeted [33–35].

References

[1] S. A. Islam, J. Luo, and M. J. E. Sternberg, "Protein Eng. Des. Sel. **8**, 513 (1995).
[2] S. J. Wheelan, A. Marchler-Bauer, and S. H. Bryant, Bioinformatics **16**, 613 (2000).
[3] F. Jacob, Science (80-.). **196**, 1161 (1977).
[4] E. Fischer, Berichte Der Dtsch. Chem. Gesellschaft **27**, 2985 (1894).
[5] D. E. Koshland, Proc. Natl. Acad. Sci. U. S. A. **44**, 98 (1958).
[6] R. D. Vale, Cell **112**, 467 (2003).
[7] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[8] A. Barth, Prog. Biophys. Mol. Biol. **74**, 141 (2000).
[9] H. D. Middendorf, R. L. Hayward, S. F. Parker, J. Bradshaw, and a Miller, Biophys. J. **69**, 660 (1995).
[10] H. Susi and D. M. Byler, *Enzyme Structure Part K* (Elsevier, 1986), pp. 290–311.
[11] Z. Ren, D. Bourgeois, J. R. Helliwell, K. Moffat, V. Šrajer, and B. L. Stoddard, J. Synchrotron Radiat. **6**, 891 (1999).
[12] F. Schotte, M. Lim, T. A. Jackson, A. V Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff, and P. A. Anfinrud, Science (80-.). **300**, 1944 (2003).
[13] A. K. Mittermaier and L. E. Kay, Trends Biochem. Sci. **34**, 601 (2009).
[14] A. Mittermaier and L. E. Kay, Science (80-.). **312**, 224 (2006).
[15] R. Sprangers, A. Gribun, P. M. Hwang, W. A. Houry, and L. E. Kay, Proc. Natl. Acad. Sci. U. S. A. **102**, 16678 (2005).
[16] K. a Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, and D. Kern, Nature **450**, 913 (2007).
[17] F. Zhang and R. Brüschweiler, J. Am. Chem. Soc. **124**, 12654 (2002).
[18] J. García de la Torre, M. L. Huertas, and B. Carrasco, J. Magn. Reson. **147**, 138 (2000).
[19] T. Heyduk, Curr. Opin. Biotechnol. **13**, 292 (2002).
[20] S. Weiss, Nat. Struct. Biol. **7**, 724 (2000).
[21] I. D. Sahu, R. M. McCarrick, and G. A. Lorigan, Biochemistry **52**, 5967 (2013).
[22] R. Graceffa, R. P. Nobrega, R. A. Barrea, S. V Kathuria, S. Chakravarthy, O. Bilsel, and T. C. Irving, J. Synchrotron Radiat. **20**, 820 (2013).
[23] R. Biehl and D. Richter, J. Phys. Condens. Matter **26**, 503103 (2014).
[24] R. K. Scopes, Enzym. **8**, 335 (1973).
[25] T. N. Bryant, H. C. Watson, and P. L. Wendell, Nature **247**, 14 (1974).
[26] B. E. Bernstein, P. A. Michels, and W. G. Hol, Nature **385**, 275 (1997).
[27] R. Inoue, R. Biehl, T. Rosenkranz, J. Fitter, M. Monkenbusch, A. Radulescu, B. Farago, and D. Richter, Biophys J **99**, 2309 (2010).
[28] G. Harauz, N. Ishiyama, C. M. D. Hill, I. R. Bates, D. S. Libich, and C. Fares, Micron **35**, 503 (2004).
[29] E. Polverini, A. Fasano, F. Zito, P. Riccio, and P. Cavatorta, Eur. Biophys. J. **28**, 351 (1999).
[30] P. Bernado, E. Mylonas, M. V Petoukhov, M. Blackledge, and D. I. Svergun, J Am Chem Soc **129**, 5656 (2007).
[31] A. M. Stadler, L. Stingaciu, A. Radulescu, O. Holderer, M. Monkenbusch, R. Biehl, and D. Richter, J. Am. Chem. Soc. **136**, 6987 (2014).
[32] K. Adachi, R. Yasuda, H. Noji, H. Itoh, Y. Harada, M. Yoshida, and K. Kinosita, Proc. Natl. Acad. Sci. U. S. A. **97**, 7243 (2000).
[33] O. Korb, T. S. G. Olsson, S. J. Bowden, R. J. Hall, M. L. Verdonk, J. W. Liebeschuetz, and J. C. Cole, J. Chem. Inf. Model. **52**, 1262 (2012).

[34] N. Andrusier, E. Mashiach, R. Nussinov, and H. J. Wolfson, Proteins **73**, 271 (2008).
[35] S.-Y. Huang and X. Zou, Proteins **66**, 399 (2007).
[36] G. Lipari and A. Szabo, J. Am. Chem. Soc. **104**, 4546 (1982).

E 4 Functional Amyloids

Martin Carballo-Pacheco, Bogdan Barz and Birgit Strodel
Institute of Complex Systems 6: Structural Biochemistry
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Functional Amyloids in Nature **3**

3 Functional Amyloids as Nanomaterials **7**

4 Modeling amyloid formation **10**

5 Conclusions and Outlook **12**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Amyloids are fibrils made by proteins in a β-strand conformation that lay perpendicular to the fibril axis [1, 2]. They are normally formed by two or more β-sheets which aggregate laterally to form the final fibrils. This structure is called cross-β because of its characteristic X-ray diffraction pattern with a reflection at 4.8 Å corresponding to the distance between β-strands in a sheet, and a second reflection at ~ 10 Å corresponding to the distance between β-sheets. The essential characteristics of amyloids are summarized in figure 1. It has been hypothesized that most proteins can form amyloids in the right environment [3], but most proteins in physiological conditions do not form amyloids. Proteins that form amyloids in physiological conditions have a high β-sheet propensity and their monomers are usually disordered.

Amyloids have been extensively studied because of their association with various diseases. In particular, the deposits found in the brains of patients with Alzheimer's or Parkinson's disease are composed of amyloids. A list of some of the diseases related to amyloids can be seen in table 1. Even though amyloids are related to many diseases, it is still unclear what role they play in these diseases. Notably, it has been observed that the concentration of fibrils does not correlate with the stage of the disease [5]. However, the concentration of soluble oligomers, made up of only a few peptides, have a stronger correlation with the stage of the disease. This

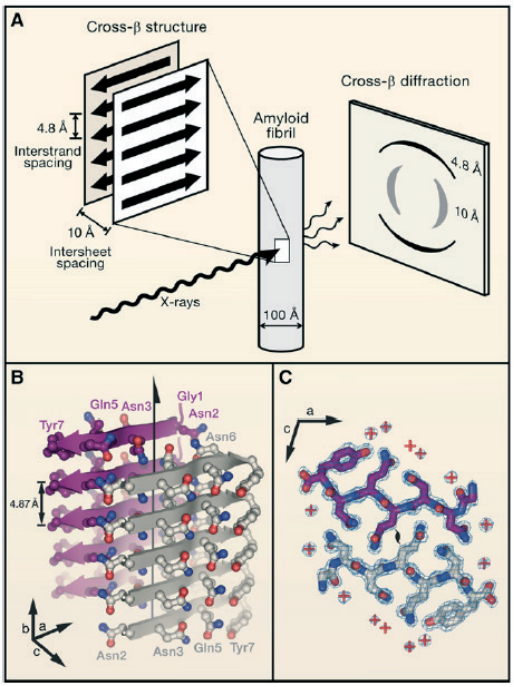


Fig. 1: *(A) Amyloids have a cross-β diffraction pattern with a reflection at 4.8 Å caused by the interstrand spacing and a second diffraction around 10 Å caused by the intersheet spacing. In (B) and (C), the structure of GNNQQNY microcrystals from Sup35, the first amyloid structure resolved, is shown [4]. Amyloids are stable because of the hydrogen bonds between the backbone atoms of the peptides. This figure is reproduced with permission of Elsevier.*

Table 1: *Disease-related amyloids.*

Protein	Disease	Reference
β-amyloid	Alzheimer's disease	Benilova et al. [7]
α-synuclein	Parkinson's disease	Irwin et al. [8]
IAPP (amylin)	Diabetes mellitus type 2	Westermack et al. [9]
Huntingtin	Huntington's disease	Perutz [10]
Prion protein PrP^{Sc}	Transmissible spongiform encephalopathy	Aguzzi and Calella [11]

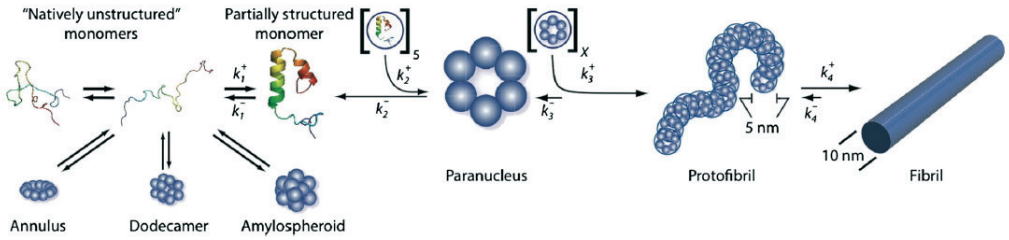


Fig. 2: *Aggregation pathway of monomers of β-amyloid to amyloid fibrils . Monomers are usually unstructured. These monomers can form off-pathways aggregates such as amylospheroids. Monomers can also form partially structured monomers which later form paranuclei on the path to amyloid fibrils. These paranuclei aggregate into protofibrils, which in turn assemble into fibrils [12]. This figure is reproduced from [12].*

has led to the hypothesis that it is not amyloid fibrils but soluble oligomers which are the toxic species [6]. It has also provoked the extensive study of amyloid aggregation, which was found to be a very complex process with many intermediate species (see figure 2 for the aggregation mechanism proposed for β-amyloid). It is also still unclear why these oligomers are the toxic species. For the moment, the hypothesis with the strongest experimental evidence is that these oligomers are toxic because of their interactions with cell membranes.

Most amyloids are associated to diseases. However, the last few years have seen the discovery of amyloids which have normal physiological roles. These amyloids are called *functional amyloids*. This lecture concentrates on these recently discovered amyloids. In particular, it covers both functional amyloids which are found in nature, and those in nanomaterials, which have been recently designed taking advantage of the amyloid fold.

2 Functional Amyloids in Nature

The discovery of functional amyloids has changed our understanding of the amyloid fold. Today, we know that functional amyloids can be found in almost every kingdom (see table 2). The functions of these amyloids are diverse and include storage of peptides, scaffold for melanin formation, adhesives, or biofilms.

The first functional amyloid was found in the fungus *Podospora anserina* in 1997 and is made of the HET-s protein [13]. *Podospora anserina* is a filamentous fungus which can have heterokaryon cells, i.e., cells with two different nuclei. Such condition occurs when two cells undergo vegetative cell fusion, which can happen within the same individual but also between

Table 2: *Functional amyloids in biology.*

Protein	Organism	Function	Reference
HET-s	*Podospora anserina* (fungus)	Heterokaryon incompatibility	Costou et al. [13]
Curli	*Escherichia coli* (bacterium)	Extracellular adhesion, invasion and biofilm formation	Chapman et al. [14]
Pmel17	*Homo sapiens*	Melanin scaffold	Fowler et al. [15]
Peptides hormones	Mammalia and Amphibia	Storage and release of hormones	Maji et al. [16]
RIP1/RIP3	*Homo sapiens*	Signalling complex for programmed necrosis	Li et al. [17]
CPEB	*Aplysia* (mollusca)	Long-term memory	Si et al. [18]

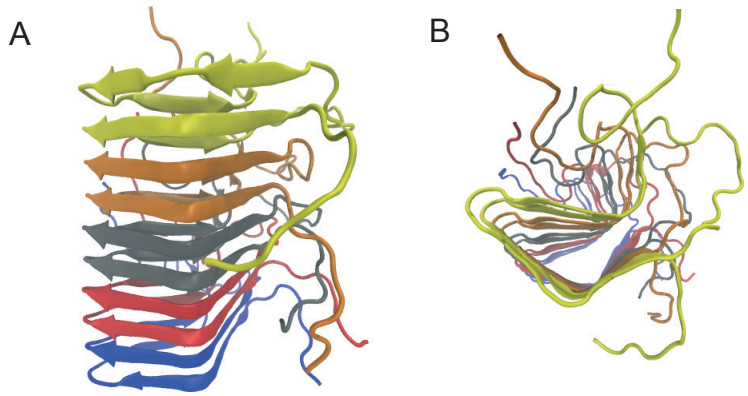


Fig. 3: *Structure of the Het-s amyloid fibrils from (A) side-view and (B) top-view [19]. Each monomer forms two windings and is represented by a different color.*

different individuals. However, to limit infections, such cell fusion must be tightly controlled. In particular, cells must check if the two nuclei are incompatible. To check such incompatibility, fungi have certain loci to control the viability of the heterokaryon cell. One such loci is the HET-s protein. If two different antagonistic alleles are expressed in the same cytoplasm, the cell is killed. Different alleles code for proteins which differ in 14 amino acids, but only one difference is enough to kill the cell. Coustou et al. [13] showed that Het-s can form heterooligomers which have an amyloid structure. These amyloids formed by proteins from different alleles cause the death of the cell. Wasmer et al. [19] determined the structure for the section 218-289 of the Het-s protein (see figure 3). As other amyloids, Het-s has a very high β-sheet content. However, instead of forming two β-sheets it forms a β-solenoid, where each protein forms two windings. Such structure is obtained from the pseudo-repeat primary structure. This structure is much more complex than that of aberrant amyloids and it resembles the complexity of folded proteins [19]. Such organization is in concordance with the fact that HET-s is not polymorphic as many aberrant amyloids are. Furthermore, it exemplifies the fact that HET-s has evolved for its role, in opposition to aberrant amyloids which are caused by misfolded proteins.

Functional amyloids were also discovered in bacteria such as *Escherichia coli* [14]. *Escherichia coli* produces different proteinaceous filaments which are used extracellularly to promote col-

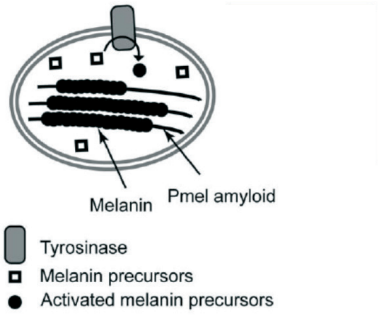


Fig. 4: *Pmel17 forms amyloids inside melanosomes. Later, melanin precursors are activated through tyrosinases. Pmel17 amyloids template the aggregation of the activated melanin precursors [15]. This figure is reproduced from [15].*

onization, enter into host cells and to organize communities such as biofilms. One of these filaments used to colonize inert surfaces are composed of fibrils called curli. These fibrils have all the properties of amyloid fibrils. Curli fibrils also have a complex network of protein interactions for their formation. In particular, the protein CsgB is needed to nucleate the aggregation of the protein CsgA, which is the main component of the curli fibrils. Other proteins such as CsgE, CsgF and CsgG are also involved. Such complex mechanism also shows the influence of evolution in the mechanism of functional amyloid formation.

The first functional amyloids found in mammals was Pmel17 [15]. Pmel17 is a protein that is essential for the production of melanosomes. Melanosomes are large organelles which store melanin, a pigment which is found in the skin and is responsible for the protection from UV rays. Melanin is formed by the aggregation of melanin precursors, which first must be activated by tyrosinases. Pmel17 forms amyloid fibrils which serve as a scaffold for such aggregation. A sketch of the process can be seen in figure 4. Considering that melanin precursors are toxic, melanin formation must be strictly regulated and Pmel17 plays a role in this regulation mechanism. Furthermore, amyloids may even be toxic themselves and they also need to be highly regulated. In fact, Pmel17 is trafficked to melanosomes as a transmembrane protein that cannot aggregate. Later, the Mα section of the protein, which is the section that aggregates, is secreted through proteolysis. This section then aggregates and forms amyloids. This process is very similar to the pathway in which Alzheimer's β-amyloid is created. One of the most interesting characteristics of Pmel17 aggregation is that it is much faster than the aggregation of toxic amyloid such as β-amyloid or α-synuclein. This is one of the possible mechanisms for amyloids to avoid being in the oligomer state and diminish their toxicity.

Furthermore, a number of peptide hormones have been found to be stored as functional amyloids [16]. These hormones are stored in the Golgi apparatus and later transported in secretory granules to the extracellular space where these amyloid disaggregate into monomers, which are the active species. This process is illustrated in figure 5. In this case, the amyloid fold is an excellent fold for such a role as hormones can be stored in a high concentration. Storing peptides in such a manner will help for a rapid secretion, as the synthesis of the peptide would take much longer. The amyloid fold can also explain how secretory granules are composed of single peptide species, as amyloid formation is sequence specific. Maji et al. [16] observed that the hormones aggregate only under certain environmental conditions, such as at a specific

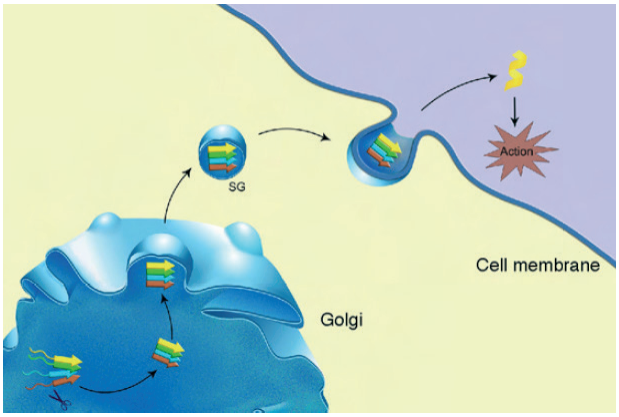


Fig. 5: *Sketch of how hormone peptides are stored as functional amyloids. Hormones are stored as amyloids in the Golgi apparatus. Later, they are transported in secretory granules to the extracellular space. When outside the cell, the amyloids dissolve and hormone monomers become active [16]. This figure is reproduced from the Supporting Material of [16].*

pH or in the presence of helper molecules such as glycosaminoglycans (GAGs). Such delicate environmental conditions could be relevant to how these functional amyloids avoid being toxic.

The last functional amyloid discovered in humans was a heterooligomeric amyloid signalling complex formed by RIP1 and RIP3 kinases [17]. These kinases are required for programmed necrosis and function in a feed forward mechanism where kinase activation and RIP1/RIP3 aggregation reinforce each other. When RIP1 and RIP3 are not phosphorylated they do not aggregate but when phosphorylated they form amyloids. Interestingly, when aberrant amyloids, such as τ-protein and α-synuclein, are phosphorylated they aggregate faster too. As these kinases are signalling for programmed necrosis, it has been hypothesised that these amyloids are the ones that kill the cell.

Finally, one of the most fascinating functionalities in which amyloids may play a role is long-term memory [18, 20]. Long-term memory, unlike short-term memory, needs the creation of new synapse pathways between neurons. The creation of new synapses is mediated by serotonin, a monoamine neurotransmitter. However, these changes must be made permanent and how this happens has been challenging to understand. Considering that amyloid fibrils are much more stable than most proteins, it has been suggested that they play a role in long-term memory. In particular, Si et al. [18, 20] showed that the cytoplasmic polyadenylation element binding protein (CPEB), which is related to long-term memory in Aplysia (a sea slug usually studied by neurobiologists because it has only around 20,000 neurons), forms amyloids. In particular, CPEB appears to be in two different states: a first one which is not active and does not aggregate, and a second one in which it is active and aggregates into a prion-like amyloid. In the prion-like state, it stimulates other non-active CPEB to become active and aggregate, and later bind to RNA. Serotonin increases the production of CPEB. There are still many open question in how amyloids relate to long-term memory. Furthermore, it would be interesting to observe how this translates into the human brain, which is much more complex than the one from Aplysia.

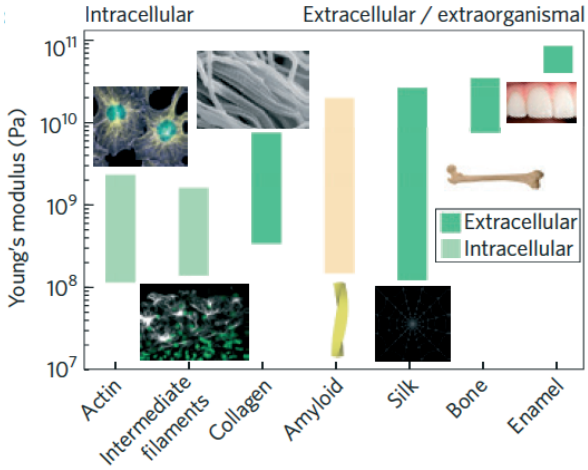


Fig. 6: *Young's moduli for different natural products. The most rigid materials are found outside the cells. Amyloids, which are found both intra- and extracellularly, can be as rigid as silk or bones [21]. This figure is reproduced with permission of Nature Publishing Group.*

3 Functional Amyloids as Nanomaterials

Since the discovery of functional amyloids, scientists and engineers have been thinking about taking advantage of the amyloid fold to create nanobiomaterials [21]. Amyloids are very stable, biocompatible (at least in their functional roles), self-aggregating and can have varied functionality. They can also be as rigid as some of the most rigid natural materials such as silk and bones (see figure 6). Moreover, they assemble from soluble precursors and could be built in a bottom-up way from simple building blocks. These properties make them very appealing for many industrial applications (see table 3). Moreover, amyloids can be combined with more traditional nanomaterials such as carbon nanotubes to produce hybrid materials [22].

Table 3: *Functional amyloids as nanomaterials.*

Function	Reference
Long-acting drugs	Maji et al. [23]
Nanowires	Baldwin et al. [24]
Nanostructured films	Fowler et al. [25]
Catalytic amyloids	Rufo et al. [26]
Carbon dioxide capture	Li et al. [27]

One of the possible applications of functional amyloids is drug delivery. Many pharmacological drugs must be administered regularly to have a lasting and durable effect. This poses no problem for drugs that can be administered orally. However, many drugs cannot withstand the acid and enzyme-rich environment of the digestive system. Therefore, these drugs are usually given through injections. However, it is not convenient to inject patients regularly, sometimes more than once a day, particularly if they are not in an in-hospital situation. Therefore, it would be of great benefit to have drugs that are injected sparingly and whose effects last long. One of the

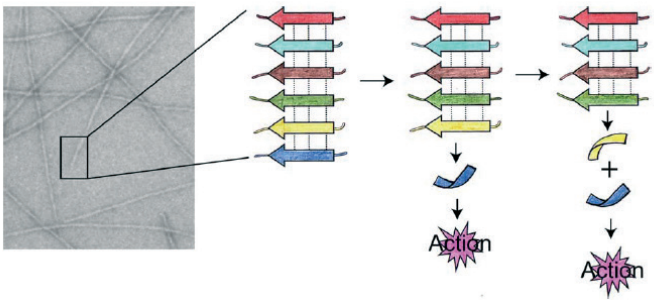


Fig. 7: *Sketch of how amyloid fibrils can be used as long-lasting drugs. On the left, an electron microscopy image is shown. The fibril is stable and monomers are released slowly. These monomers are the active species [23]. This figure is reproduced from [23].*

main characteristic of the amyloid fold is that it is more stable than folded soluble proteins and may provide a possible solution to this problem. In fact, it has been shown that gonadotropin-releasing hormones, which are used to treat hormone-related cancers such as prostate cancer, can form amyloids which have long lasting effects [23]. In this test, peptides amyloids are administered and later slowly dissociate into monomers, which are the active species. A sketch of the mechanism can be found in figure 7.

Another advantageous property of amyloid peptides is that they can self-assemble at a nanoscale. Thus, one could use an amyloidogenic peptide as template for a material that does not assemble on its own. This has been used for the production of nanowires [24]. For this purpose, a protein composed of two linked sections must be synthesized. The first section is an amyloidogenic peptide such as the tandem repeat of the SH3 domain. The second section is a metalloprotein such as cytochromes. Cytochromes are proteins which have a heme group (i.e., they contain Fe^{2+}) and are used in nature for electron transport. By binding an amyloid forming domain with a domain that allows for electron transfer, and aligning the electron transfer domain correctly, an electrical conducting nanowire is built.

Functional amyloids have also been used is in the context of organic photovoltaics. When designing photovoltaic materials, one of the most relevant properties is the surface between electron donors and acceptors. In this case, amyloids can be used as templates, increase the donor-acceptor interface and improve electron transport [28]. This was done by incorporating amyloid nanowires into the active layer of the organic solar cell. These structures can be employed for light-harvesting purposes.

Another possible nanotechnological application of amyloid fibrils is the formation of nanostructured films [25]. In this application, films were produced in a two-step process: in a first step, amyloid fibrils are formed under aggregation-favorable conditions, and in a second step, the fibrils are cast in a thin film, in which the fibrils align in the film plane. In such a film, the self-assembly follows a hierarchy of scales: within the fibrils there is a nanometer-scale ordering and between fibrils there is a micrometer-scale ordering. Using this method, films of up to a few centimetres could be manufactured. Moreover, the films have an elastic module in the order of tens of GPa (see figure 6), which is similar to the mechanical properties of individual fibrils, and also, similar to the strongest protein materials such as collagen. The high transferability of properties from single fibrils to a film is essential, as one of the most common problems in the

production of macroscopic materials starting from nanomaterials is the lack of transferability of properties between scales. This is, for example, the case for carbon nanotubes where the lack of transferability is caused by the lack of robust contacts between different carbon nanotubes. Amyloid fibrils, on the other hand, interact via side chains, which makes such transferability of properties possible. These amyloid films were later used to create a surface of fluorophores, which interact with the amyloid proteins. In general, the possible applications for such amyloid films are vast.

Most nanotechnological functional amyloids take advantage of the self-assembly and strength of the amyloid fold. However, considering that amyloids can have functional roles on their own, it can be hypothesized that amyloids themselves may play direct functional roles. A very interesting advance in that field is the design of catalytic amyloid fibrils by Rufo et al [26]. They designed peptides that forms amyloids and mimic the active site of carbonic anhydrase. In particular, the active site of carbonic anhydrase has three histidines: two of them in a β-strand at positions i and $i+2$ and the third one in a neighbouring β-strand. Such motif binds to Zn^{2+} which lowers the pK_a of water, stabilizing hydroxide, which later suffers a nucleophilic attack from carbon dioxide to produce bicarbonate. Having this in mind, peptides which self-assemble into amyloids and produce the histidine triangle motif where synthesised (e.g. IHIHIQI). The structure of the amyloids is shown in figure 8. These peptides were successfully tested for catalyzing the conversion of carbon dioxide and water into bicarbonate. These enzymatic amyloids could have nanotechnological applications and might have played a role in the beginning of life. Most enzymes are over 100 amino acids long and, therefore, it is hard to explain how such large proteins existed in the first living beings. Thus, enzymatic amyloids, which are much shorter, could have been a first step in the evolution of modern enzymes.

Finally, functional amyloids have been used for carbon dioxide capture [27]. Carbon dioxide is the main greenhouse gas and it has an enormous contribution to man-made climate change. One of the possible solutions to decrease the production of industrial carbon dioxide is the use of materials which capture carbon dioxide in the flue gases of industries and power plants. The most promising method for carbon dioxide capture is the use of monoethanolamine (MEA). However, MEA is toxic, flammable, corrosive and volatile. Moreover, up to 30% of the energy production has to be used to regenerate it. Therefore, the market is waiting for new materials for carbon dioxide capture. Normally, carbon dioxide capture is realized by a reaction of primary alkanolamine to carbamate through the following reaction:

$$\mathrm{RNH_2 + CO_2 \longrightarrow RNH_2^+COO^-}, \tag{1}$$

$$\mathrm{RNH_2^+COO^- + B \longrightarrow RNHCOO^- + BH^+}, \tag{2}$$

where B represents a base, which is usually a second amine. Having this in mind, Li et al. [27] designed a peptide which aggregates into amyloids and includes a lysine amino acid, which has an amine group. The peptide they used was VQIVYK, which is a fragment of the τ-protein and forms amyloid fibrils. At high pH, the amino group of the lysine is uncharged which makes it available for carbamate formation. Fibrils of VQIVYK were produced, centrifuged and lyophilized to produce a white powder. They showed that this peptide material can adsorb carbon dioxide in a 1:1 stoichiometry, which means that, in this case, the bases are probably buffer salts. The fibrils can be regenerated when heated up to a 100°C. Finally, fibrils can also adsorb carbon dioxide even when the gas has a high percentage of water vapor. Even though amyloid fibrils may have a possible future in carbon dioxide capture, there are still many hurdles to pass. For example, the current amyloid materials cannot work in the acidic and high

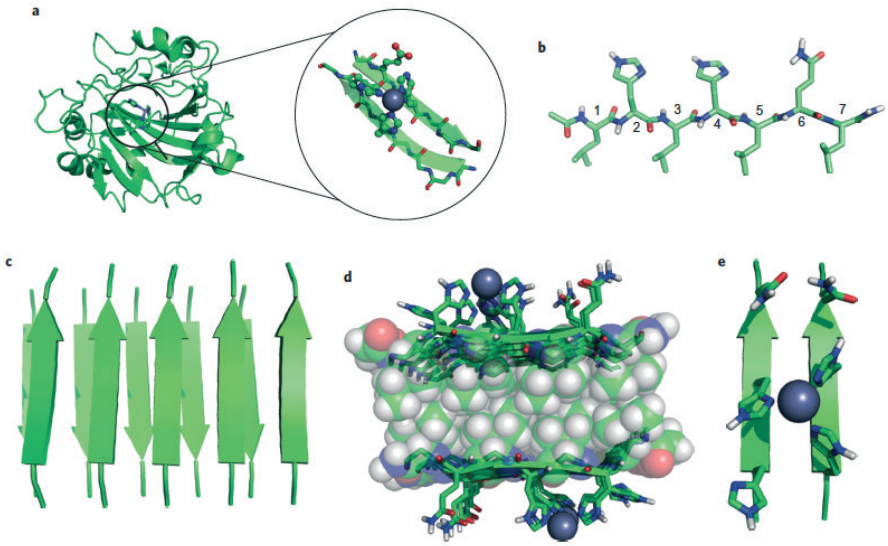


Fig. 8: *Design of catalytic amyloid fibrils. (a) The structure of human carbonic anhydrase with a zinc binding motif. (b) Structure of the IHIHIQI monomer in an extended conformation. (c,d) Structure of the catalytic amyloid fibrils from top and side. (e) The zinc binding domain from carbonic anhydrase is replicated in the amyloid fibrils [26]. This figure is reproduced with permission of Nature Publishing Group.*

temperature conditions of standard flue gas, and amyloids have lower adsorption capacity than MEA. However, amyloids have many possible advantages such as mild regeneration temperature, compatibility with water, self-assembly and biodegradability. It is thought that in the future mass bacterial production of amyloid peptides could also produce savings in mass production.

4 Modeling amyloid formation

In the following it is described for the Alzheimer's Aβ peptide how the formation of amyloid oligomers depicted in figure 2 can be assessed via molecular simulations. The modeling of functional amyloids can be performed in a similar manner and is currently underway in the Strodel group. Oligomers of amyloid-β protein (Aβ) are considered one of the main causes of neurotoxicity and are thus highly associated with the onset of Alzheimer's disease [29]. Experimental methods are able to identify some characteristics of aggregating proteins such as the oligomer size distribution or cellular toxicity [30], but due to the fast conversion of oligomers into fibrils, the elucidation of their structure at molecular level is challenging. Computational methods have the advantage of atomistic detail but are exposed to the challenge that the size and time scales in experiment are usually larger than in atomistic simulations. In a recent study we reported the early assembly of Aβ_{1-42} (Aβ42 for short) proteins at experimental concentrations using all-atom molecular dynamics (MD) simulations in implicit solvent, which were initiated from 20 isolated Aβ42 monomers [31]. To describe the assembly process we derived a maximum flow transition network (MTN) based on aggregation states defined by $N1|N2|N3$, where $N1$ represents the oligomeric size, $N2$ is the average number of hydrogen-bonds between in-

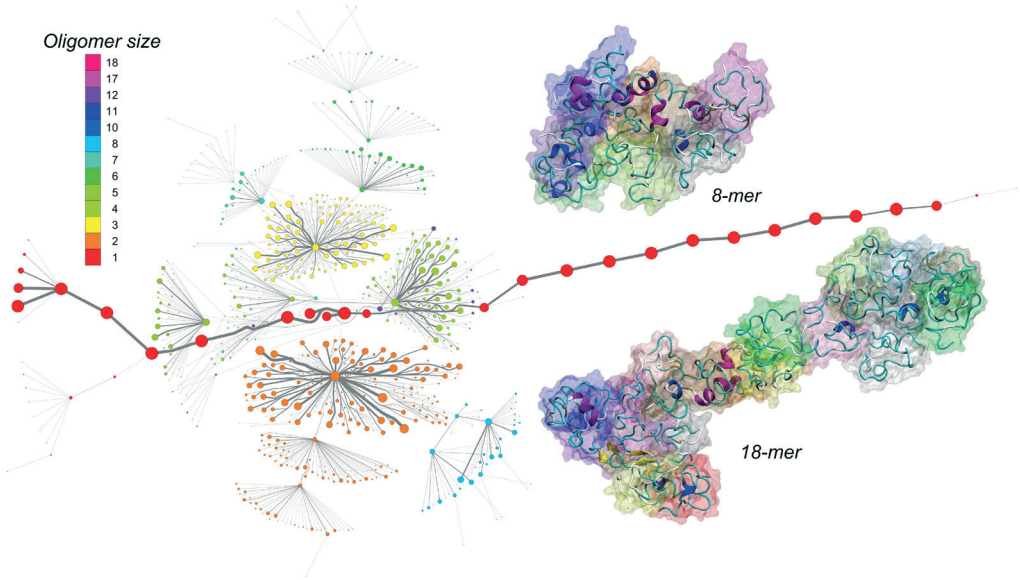


Fig. 9: *Maximum flow transition network. The node area is proportional to the state population and the color corresponds to the oligomer size. The edge thickness correlates with the number of transitions. Two representative oligomer structures for the octamer and 18-mer are shown in cartoon and surface representation.*

dividual chains from the oligomer, and $N3$ is the average number of amino acids in β-strand conformation per peptide in the oligomer. Detailed information about the methods are provided in [31, 32].

The MTN in figure 9 displays a complex aggregation process in which initial monomers assemble into oligomers up to 18-mers during 200 ns. The aggregation states with $N1 = 1$ are distributed linearly with a gradual increase in β-strand content from right to left. The monomers with more β-strands more readily aggregate than the others: state $1|0|17$ is the main connection node to oligomers $2|2|15$, $3|4|16$, $4|6|13$, $5|6|17$, $12|7|13$ and $17|8|13$. These states are also the central connection to the other states with the same $N1$ value. In addition, state $1|0|20$ is in direct contact with state $18|9|13$ and thus to the rest of the 18-mer cluster. A representative snapshot of the 18-mer at the end of the simulation indicates an elongated conformation rather than a globular one as observed for the 8-mer (figure 9). The central aggregation states for hexamer and heptamers are $6|9|15$ and $7|7|16$, respectively, which are preferentially formed from trimers. Dimers are directly connected to aggregation state $8|7|12$ and thus to the other $N1 = 8$ states, indicating that octamers are largely formed by the addition of dimers to either tetramers or hexamers. We calculated the oligomer mass distribution, which revealed a higher population for dimers, tetramers, hexamers, octamers, 12-mers and 18-mers. Previous computational [33] and experimental studies [30] of Aβ aggregation report significant peaks in pentamers/hexamers and 12/13-mers, in agreement with our results.

To describe structural changes during the assembly process we calculated the time evolution of the secondary structure propensities. The initial β-strand propensity ($\sim$10%) decreased slightly throughout the simulation and had an average of 7.6$\pm$2.2%. The average helical propensity was around 10.1$\pm$4.0% and the average coil propensity started around 80% and increased slightly,

having an average value of 82.2±4.3%. This indicates little change in the overall secondary structure. Experimental studies [30] indicated 13–20% β-sheet for Aβ42 and Aβ40 while a more recent study shows a jump from ∼25% to ∼45% when converting from monomers to tetramers of Aβ40 [34]. Another experimental study reported Aβ42 oligomers without β-sheet structure, in close agreement with our findings, which are on-pathway intermediates for fibril formation [35].

To investigate the key amino acids involved in the assembly process we calculated contact maps for the interface between any two proteins that are part of an oligomer [31]. An important result from the inter-molecular contact map is the proximity of hydrophobic regions from the C-terminus including L31–A42. In addition, region L33–V35 is in close proximity to region L17–F20. On the basis of solid-state NMR spectroscopy it has been shown that the C-terminal region is buried inside disc-shaped oligomers (pentamers and 10-mers) with strong contacts between F19 and L34 [35]. The 8-mer shown in Fig. 9 matches the description of the low-order oligomers with non-β-sheet secondary structure observed by Ahmed et al. [35]. Electrostatic interactions also contribute to oligomer formation by strong contacts formed by amino acid K16 with region S8–Q15 containing the negatively charged residue E11, and by K28 with region D1–D7 containing the negatively charged residues D1, E3, and D7.

Our results suggest that early Aβ aggregation precedes β-sheet formation. This is an important observation considering the low Aβ concentration in our study (∼0.8 mM) is close to experimental concentrations (∼30 μM–1 mM) [30] and smaller than concentrations from other computational studies (3.4 mM) [33]. At the same time, the structurally stable oligomers and the small changes in their secondary structure indicate a fast assembly process relative to the time scale corresponding to their conformational reorganization. Whether the oligomers observed here during a very early aggregation stage are similar to the toxic species observed in experiments is difficult to assess. While we found similarities with experimentally observed Aβ42 oligomers in terms of size and structure [35, 33], there is still a large debate regarding which Aβ oligomers are the toxic ones [36]. Some groups consider small oligomers with high content of β-sheet as toxic species [34], others suggest a second nucleation process where amyloid fibrils are present with small oligomer species as the source for toxic oligomers [37], and others propose that toxic Aβ oligomers have cross-β structure [38]. In future studies we plan to follow the further growth and structural conversion of the early oligomers and study their interactions with membranes. The aim should be that experiments probe at the same time the size, secondary structure and toxicity of low-order oligomers, allowing to directly relate simulation and experimental results.

5 Conclusions and Outlook

In this lecture, the aggregation of functional amyloids was discussed and explained, how this aggregation process can be elucidated by means of simulation. Amyloids have been extensively studied because of their association with diseases such as Alzheimer's disease, Parkinson's disease or type II diabetes. However, in the last years many new functional amyloids have been discovered. In particular, nature has evolved functional amyloids for various roles such as biofilm formation in bacteria, heterokaryon incompatibility testing in fungi, as a scaffold for melanin formation in mammals, as storage for hormone peptides, as a signalling complex for programmed necrosis, or even as an essential step in long-term memory formation. Considering the many advantages that amyloids have, such as self-assembly, high stability, multifunction-

ality and biocompatibility, it has been predicted that they could be used as nanomaterials. In particular, they have been used in long-acting drugs delivery, to create nanowires which can be used in organic solar cells, for nanostructured films, as synthetic enzymes or for carbon dioxide capture.

However, the relevance of amyloids in nature could even go further than what we have discussed so far. It has been suggested that amyloids could be involved in the origin of life and in the first living beings [39, 40]. Proteins are complex macromolecules that must have evolved from much simpler precursors. In particular, proteins form different folds which most probably arose from one primitive fold. One of the most interesting hypotheses is that the first fold was the amyloid fold [39, 40]. Considering that the production of the first proteins must have had low fidelity, the first fold must have been simple. Amyloids, which are only a few amino acids long, have that property. For example, there are catalytic functional amyloids which are seven amino acids long [26] (for comparison, the shortest modern enzyme has 62 amino acids). Moreover, the original fold must have had high tolerance to mutations because of this low fidelity. It would also be advantageous if the fold would be multifunctional or the function could be changed with only few mutations. Amyloids have all these properties. Moreover, amyloids are very stable and can resist harsh conditions such as extreme pH, UV radiation or high temperature. If this hypothesis is true, the amyloid was an old fold which was later replaced for other folds when living beings became more complex. It is interesting that the β-sheet aggregation propensity is inversely correlated with organism complexity [41].

In conclusion, amyloids, which have mostly been studied because of their association with diseases, can have functional roles both in nature and as man-made technological applications. Moreover, they could have even played a role in the beginning of life. All this makes amyloid a fold that should not only be considered as toxic, but as functional and essential in the functioning of living beings.

References

[1] F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**(1), 333 (2006).

[2] J. Greenwald and R. Riek, Structure **18**(10), 1244 (2010).

[3] L. Goldschmidt, P. K. Teng, R. Riek, and D. Eisenberg, Proc. Natl. Acad. Sci. U.S.A. **107**(8), 3487 (2010).

[4] D. Eisenberg and M. Jucker, Cell **148**(6), 1188 (2012).

[5] C. Schmitz, B. P. Rutten, A. Pielen, S. Schäfer, O. Wirths, G. Tremp, C. Czech, V. Blanchard, G. Multhaup, P. Rezaie, H. Korr, H. W. Steinbusch, *et al.*, Am. J. Pathol. **164**(4), 1495 (2004).

[6] W. Klein, W. Stine, and D. Teplow, Neurobiol. Aging **25**(5), 569 (2004).

[7] I. Benilova, E. Karran, and B. De Strooper, Nat. Neurosci. **15**(3), 349 (2012).

[8] D. J. Irwin, V. M. Y. Lee, and J. Q. Trojanowski, Nat. Rev. Neurosci. **14**(9), 626 (2013).

[9] P. Westermark, A. Andersson, and G. T. Westermark, Physiol. Rev. **91**(3), 795 (2011).

[10] M. F. Perutz, Trends Biochem. Sci. **24**(2), 58 (1999).

[11] A. Aguzzi and A. M. Calella, Physiol. Rev. **89**(4), 1105 (2009).

[12] D. Teplow, Alzheimers Res. Ther. **5**(4), 39 (2013).

[13] V. Coustou, C. Deleu, S. Saupe, and J. Begueret, Proc. Natl. Acad. Sci. U.S.A. **94**(18), 9773 (1997).

[14] M. R. Chapman, L. S. Robinson, J. S. Pinkner, R. Roth, J. Heuser, M. Hammar, S. Normark, and S. J. Hultgren, Science **295**(5556), 851 (2002).

[15] D. M. Fowler, A. V. Koulov, C. Alory-Jost, M. S. Marks, W. E. Balch, and J. W. Kelly, PLoS Biol. **4**(1), e6 (2005).

[16] S. K. Maji, M. H. Perrin, M. R. Sawaya, S. Jessberger, K. Vadodaria, R. A. Rissman, P. S. Singru, K. P. R. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, *et al.*, Science **325**(5938), 328 (2009).

[17] J. Li, T. McQuade, A. B. Siemer, J. Napetschnig, K. Moriwaki, Y.-S. Hsiao, E. Damko, D. Moquin, T. Walz, A. McDermott, F. K.-M. Chan, and H. Wu, Cell **150**(2), 339 (2012).

[18] K. Si, S. Lindquist, and E. R. Kandel, Cell **115**(7), 879 (2003).

[19] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, and B. H. Meier, Science **319**(5869), 1523 (2008).

[20] K. Si, Y.-B. Choi, E. White-Grindley, A. Majumdar, and E. R. Kandel, Cell **140**(3), 421 (2010).

[21] T. P. J. Knowles and M. J. Buehler, Nat. Nanotechnol. **6**(8), 469 (2011).

[22] C. Li and R. Mezzenga, Nanoscale **5**(14), 6207 (2013).

[23] S. K. Maji, D. Schubert, C. Rivier, S. Lee, J. E. Rivier, and R. Riek, PLoS Biol. **6**(2), e17 (2008).

[24] A. J. Baldwin, R. Bader, J. Christodoulou, C. E. MacPhee, C. M. Dobson, and P. D. Barker, J. Am. Chem. Soc. **128**(7), 2162 (2006).

[25] T. P. J. Knowles, T. W. Oppenheim, A. K. Buell, D. Y. Chirgadze, and M. E. Welland, Nat. Nanotechnol. **5**(3), 204 (2010).

[26] C. M. Rufo, Y. S. Moroz, O. V. Moroz, J. Stöhr, T. A. Smith, X. Hu, W. F. DeGrado, and I. V. Korendovych, Nat. Chem. **6**(4), 303 (2014).

[27] D. Li, H. Furukawa, H. Deng, C. Liu, O. M. Yaghi, and D. S. Eisenberg, Proc. Natl. Acad. Sci. U.S.A. **111**(1), 191 (2014).

[28] S. Barrau, F. Zhang, A. Herland, W. Mammo, M. R. Andersson, and O. Inganäs, Appl. Phys. Lett. **93**(2), 023307 (2008).

[29] W. L. Klein, G. A. Krafft, and C. E. Finch, Trends Neurosci **24**(4), 219 (2001).

[30] G. Bitan, M. D. Kirkitadze, A. Lomakin, S. S. Vollers, G. B. Benedek, and D. B. Teplow, Proc. Natl. Acad. Sci. **100**(1), 330 (2003).

[31] B. Barz, O. O. Olubiyi, and B. Strodel, Chem. Commun. **50**(40), 5373 (2014).

[32] B. Barz, D. J. Wales, and B. Strodel, J Phys Chem B **118**(4), 1003 (2014).

[33] B. Urbanc, M. Betnel, L. Cruz, G. Bitan, and D. B. Teplow, J. Am. Chem. Soc. **132**(12), 4266 (2010).

[34] K. Ono, M. M. Condron, and D. B. Teplow, Proc. Natl. Acad. Sci. U.S.A. **106**(35), 14745 (2009), ISSN 0027-8424, 1091-6490.

[35] M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. I. Elliott, W. E. Van Nostrand, and S. O. Smith, Nat Struct Mol Biol **17**(5), 561 (2010).

[36] I. Benilova, E. Karran, and B. De Strooper, Nat. Neurosci. **15**(3), 349 (2012).

[37] S. I. A. Cohen, S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, Proc. Natl. Acad. Sci. U.S.A. **110**(24), 9758 (2013).

[38] J. C. Stroud, C. Liu, P. K. Teng, and D. Eisenberg, Proc. Natl. Acad. Sci. U.S.A. **109**(20), 7717 (2012).

[39] J. Greenwald and R. Riek, J. Mol. Biol. **421**(4–5), 417 (2012).

[40] O. Carny and E. Gazit, FASEB J. **19**(9), 1051 (2005).

[41] G. G. Tartaglia, R. Pellarin, A. Cavalli, and A. Caflisch, Protein Sci. **14**(10), 2735 (2005).

E 5 Protein Structure

O. H. Weiergräber and G. F. Schröder
Institute of Complex Systems, ICS-6
Forschungszentrum Jülich GmbH

Contents

1 Introduction 2

2 General Architecture of Proteins 2
2.1 Structure and properties of amino acids 2
2.2 Three-dimensional Fold 4
2.3 Post-translational modifications 10

3 Structure Determination 11
3.1 Protein Crystallography 12
3.2 Phasing 14
3.3 Model Building 15
3.4 Refinement 15

References 17

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

Proteins are key components in virtually all biological processes and constitute the predominant class of biopolymers in all living cells. They exert a remarkable variety of functions in gene expression, cellular metabolism, transport and storage as well as signal transduction and mechanical support of cells and tissues. Therefore, it is not surprising that proteins are at the core of most pathogenic mechanisms leading to human diseases. This text is intended to provide a basic foundation of protein structure and function, essentially adopting a structural biologist's point of view. After a general introduction to the architecture of proteins and their assemblies, a brief overview of the most important structure determination method, X-ray crystallography, will be given.

2 General Architecture of Proteins

2.1 Structure and properties of amino acids

Proteins are polypeptides, i.e. linear polymers built from 2-amino carboxylic acids (amino acids for short); although hundreds of different amino acids have been identified in organisms, only a limited set of 20, the **proteinogenic amino acids**, are used in canonical protein synthesis (**Fig. 1**). Each type of protein is characterized by a unique sequence of amino acids, its **primary structure**, which is decoded from the respective gene by the universal and highly conserved transcription/translation machinery. While most naturally occurring polypeptide chains are between 100 and 1000 amino acids in length, a few structural proteins can contain as many as 25,000 residues.

In most of the 20 amino acids, the sp^3-hybridized α-carbon carries a carboxyl moiety, an amino group and a hydrogen, as well as a side chain of different length and character. With its four different substituents, it thus constitutes an asymmetric center, allowing for two different configurations (D and L, according to Fischer's convention). The exception is glycine, with a second hydrogen bound to the α-carbon, which is therefore achiral. The structure of proline is remarkable since its side chain is cyclized to the nitrogen, resulting in a secondary amine. While polypeptides containing D and L amino acids can be generated in vitro, natural protein synthesis is strictly stereospecific; it utilizes exclusively L enantiomers.

Formally, generation of a polypeptide chain from single amino acids, which is performed by the ribosomal machinery, is a condensation reaction, i.e. it involves removal of water. In the resulting polymer, each amino acid contributes three atoms (nitrogen, α-carbon and carbonyl carbon) to the continuous main chain. The residues in a polypeptide are usually numbered from N- terminus to C-terminus, corresponding to the order of synthesis. The peptide group has several remarkable properties, each with important implications for the three-dimensional structure of proteins. First, the amide nitrogen carries a hydrogen atom with a positive partial charge, while the oxygen bound to the carbonyl carbon is negatively polarized. Therefore, these moieties are excellent hydrogen bonding partners, which causes certain patterns in the local conformation of the chain to be strongly favored. Second, the π-electrons of the carbonyl group are delocalized over the neighboring N-C bond; as a result, the latter partially acquires double bond properties, as evidenced by its reduced bond length and a loss of free rotability. Indeed, the atoms of the CO-NH group together with the adjacent α-carbons are ideally located in a plane, which essentially restricts flexibility of the main chain to rotation of the peptide planes with respect to each other, i.e. among the N-Cα and Cα-C bonds. The corresponding **dihedral angles** are termed φ and ψ, respectively (Figure 2). A third angle, ω, defines the torsion about the amide bond.

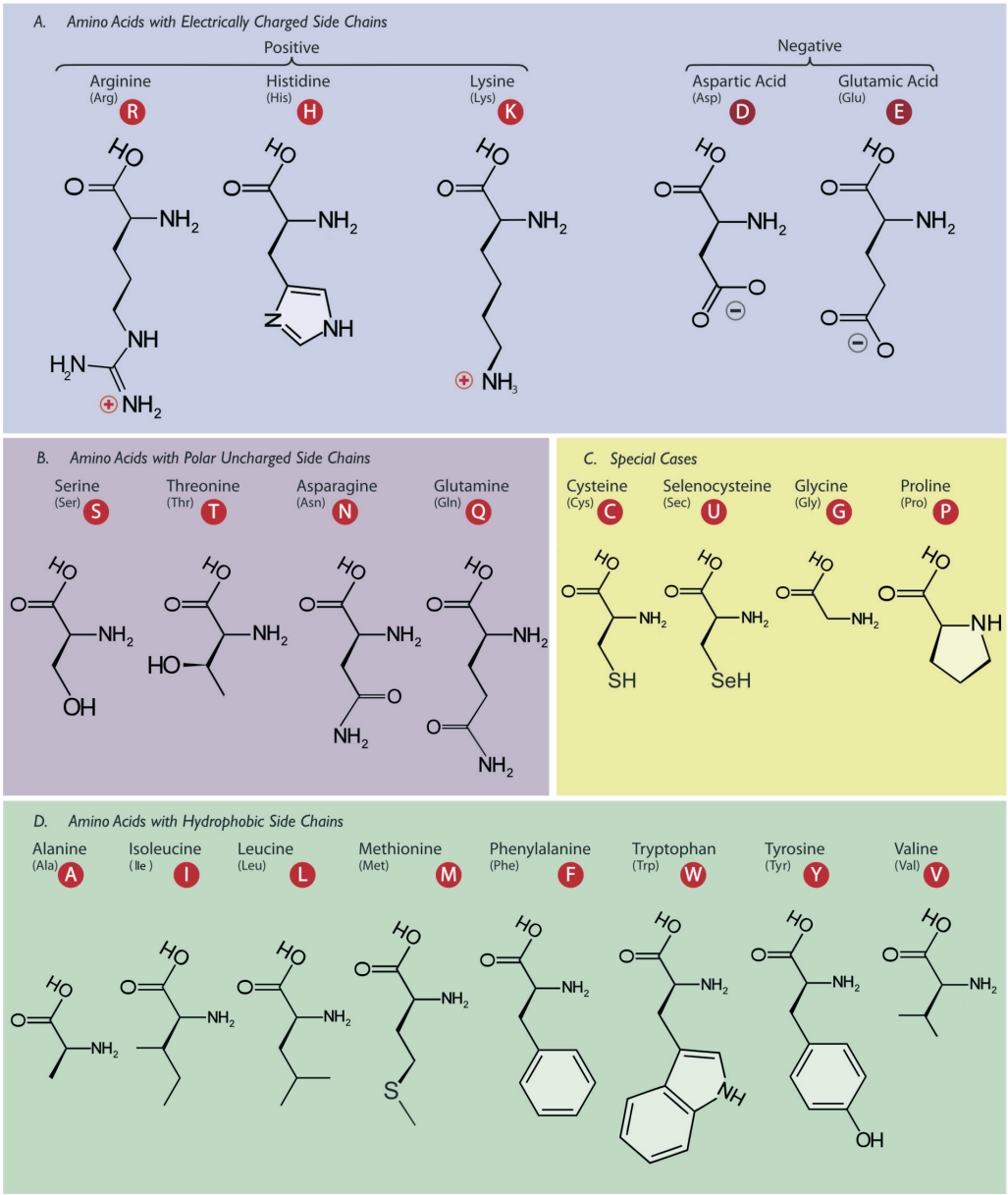


Fig. 1: *The standard amino acids found in proteins, sorted by side chain properties. Selenocysteine, which occurs in several enzymes, is incorporated via a non-standard mechanism; it is encoded by UGA (normally a stop codon) provided that specific insertion sequences are present in the mRNA. As in this chart, it is sometimes classified as a proteinogenic amino acid, raising their number to 21. Similarly, some prokaryotes are able to recruit pyrrolysine by recoding of the stop codon UAG. (Adapted from http://en.wikipedia.org/wiki/Amino acid.)*

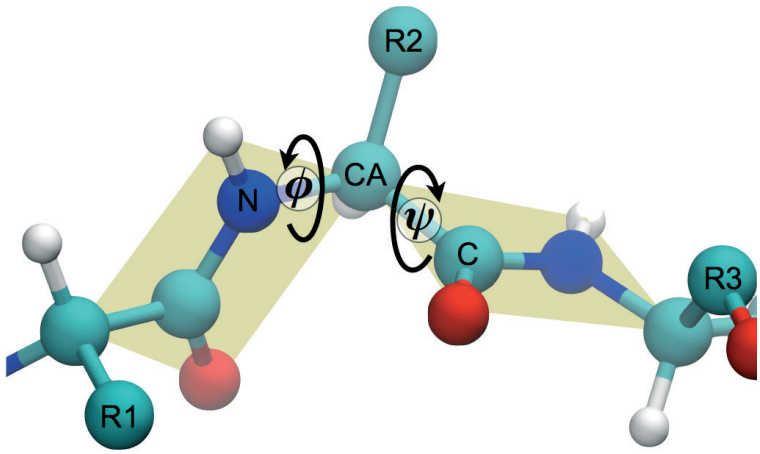


Figure 2: *The main chain conformation of a polypeptide is defined by dihedral angles* φ *and* ψ*, which determine the relative orientation of adjacent peptide planes.*

Usually, the *trans* configuration (ω approx. 180º) is strongly preferred, the exception being prolyl peptides, which are often found in *cis* (ω approx. 0º). As indicated above, the proteinogenic amino acids essentially differ by their side chains. These can be aliphatic or aromatic, and often contain one or more heteroatoms, enabling them to engage in different types of non-covalent interactions. Nonpolar residues are usually involved in hydrophobic or van der Waals contacts, while polar groups can form hydrogen bonds or, if charged, salt bridges. While all proteins are initially synthesized from this standard set of amino acids, they are often subject to covalent modifications during or after translation (discussed below).

2.2 Three-dimensional Fold

This limited repertoire of building blocks is sufficient to generate hundreds of thousands of different proteins, which are, first of all, distinguished by their unique sequence of amino acids. The great versatility of this class of biopolymers, however, mainly arises from their three-dimensional arrangement: polypeptide chains in aqueous solutions tend to collapse into compact structures, with limited thermal fluctuations around the mean positions.

Within the context of the cell, the three-dimensional structure of a mature polypeptide is determined ex- clusively by its amino acid sequence (primary structure). This has first been demonstrated by Christian B. Anfinsen, who performed pioneering experiments on denaturation and refolding of ribonuclease A [1]. Therefore, all aspects of folding (commonly classified in a hierarchical system) must ultimately relate to the properties of amino acid side chains.

Folding of a polypeptide chain into a compact three-dimensional structure is mostly promoted by the differential interaction of its residues with the solvent. Hydrophobic side chains do not interact favorably with water, which is a polar solvent. As a result, water molecules highly ordered hydrogen bonding network in the vicinity of apolar moieties, which reduces its diffusion and thus its entropy.

Table 1: *Hydrogen bonds found in proteins*

Type of hydrogen bond		Donor-acceptor distance [Å]	Reduction from v.d.W. distance
hydroxyl-hydroxyl	O–H···O	2.8 ± 0.1	25 %
hydroxyl-carbonyl	O–H···O	2.8 ± 0.1	25 %
amide-carbonyl	N–H···O	2.9 ± 0.1	20 %
amide-hydroxyl	N–H···O	2.9 ± 0.1	20 %
amide-imidazole	N–H···N	3.1 ± 0.2	15 %
amide-sulfur	N–H···S	3.7	10 %

According to the Gibbs equation (G = H − TS), this generates a positive (i.e. unfavorable) contribution to the overall free energy, which is minimized by the system. In an aqueous environment, apolar residues therefore tend to be sequestered within the core of the molecule, avoiding contact with water, whereas charged and polar side chains are usually exposed on the surface. The presence of polar moieties in the interior of a protein cannot be avoided completely; in particular, this is true for the polarized N-H and C=O bonds, which are part of every peptide linkage. In order to obtain a stable fold, the hydrogen bonding potential of these groups needs to be satisfied to the maximum extent possible. Table 1 lists the types of hydrogen bonds commonly observed in protein structures. By definition, hydrogen bond donor and acceptor are the non-hydrogen atoms involved, with the former carrying the hydrogen atom and the latter contributing a free pair of electrons. The spread in donor-acceptor distances mainly arises from the variable angle defined by the three atoms. Formation of hydrogen bonds between backbone atoms leads to the establishment of two important types of secondary structure, α-helices and β-strands, while side chain interactions are crucial for the packing of these elements in order to achieve a compact fold. The mechanisms of protein folding will be the topic of a subsequent lecture; therefore, the following paragraphs will focus on the outcome of the process, i.e. structural features regularly found in mature polypeptides.

Secondary structure

The term secondary structure refers to certain patterns of backbone torsion angles within stretches of adjacent amino acid residues, which are usually engaged in a regular and well-defined set of hydrogen bonds. In the **α-helix**, the polypeptide backbone is folded into a right-handed spiral, with about 3.6 residues per turn. This arrangement allows for the formation of a favorable hydrogen bond between the carbonyl oxygen of residue n and the amide nitrogen of residue n+4. Since this hydrogen bond effectively closes a ring of 13 atoms, the α-helix is alternatively named 3.6_{13} helix. The interior of the helix is densely populated with atoms, leaving virtually no unoccupied space. A catchy mnemonic for this arrangement is the ”Christmas tree” analogy: In vertical orientation with the N-terminus at the bottom, the side chains point slightly down like the branches of the tree, while the carbonyl groups are the candles standing upright (Figure 3).

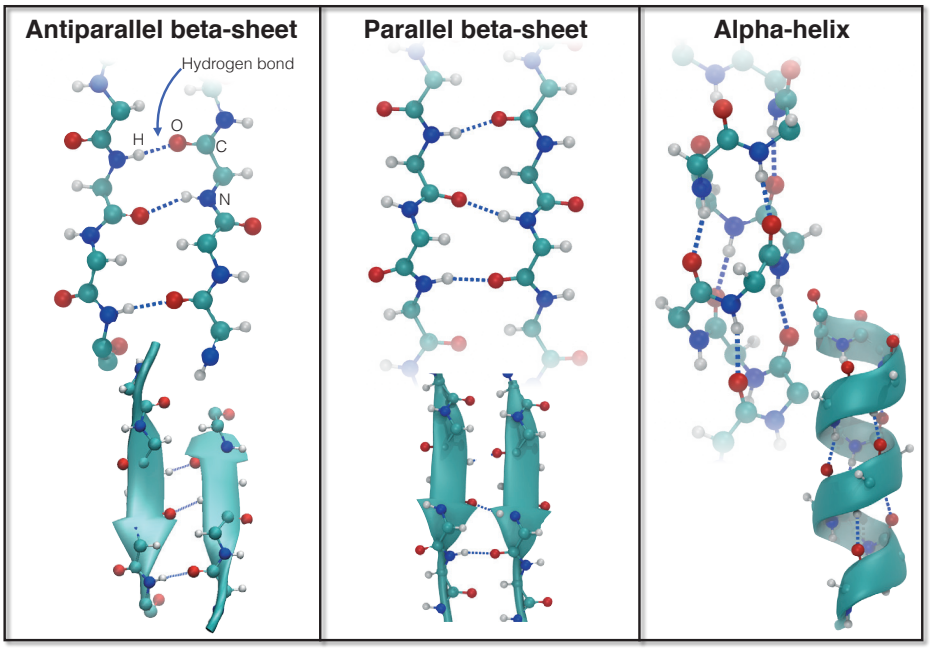


Figure 3: *Major secondary structure elements found in proteins.*

Obviously, the regular hydrogen bonding pattern will not hold for the four residues located at each terminus of an α-**helix**; these stretches can be stabilized by interaction with suitable side chain atoms. The almost parallel orientation of the polarized hydrogen bonding groups entails a dipole moment for the entire helix, which will obviously grow with the length of the structure. This is the main reason why very long α-helices (exceeding approx. 40 residues) become unstable; in fact they will tend to collapse into a helix-turn-helix structure, resulting in an antiparallel arrangement of dipoles.

Helical structures with different characteristics have been found, but since these are energetically less favorable, they are usually confined to short stretches of amino acids (**Table 2**). Specifically, 3_{10}-helices are often found within loop regions, but also occur at the termini of regular α-helices. The latter is also true for π-helices (4.4_{16}-helices). Due to the unique stereochemical properties of glycine and proline, polymers of these amino acids allow for additional helical conformations. In contrast to the structures introduced above, polyproline type II forms a left-handed helix; it is known as the target motif for SH3 domains[1] and is thus implicated in protein-protein recognition. The polyglycine type II helix is closely related to polyproline type II, as far as backbone torsion angles are concerned. Note that the corresponding type I forms are not considered here. The polyproline type I helix is right-handed, but has all peptide bonds in *cis* configuration, which makes it very unlikely to occur in natural proteins; polyglycine type I does not have a helical conformation.

[1] The SH3 (Src-homology 3) domain is a small module of about 60 amino acids found in many proteins involved in signal transduction.

Table 2: *Geometric properties of helical protein conformations*

Helix Type	Frequency	H-bonding	Handedness	Typical ϕ	Typical ψ
α-helix(3.16_{13})	abundant	i···i+4	right	-57º	-47º
3_{10}-helix	infrequent	i···i+3	right	-20º	-54º
π-helix	rare	i···i+5	right	-57º	-80º
polyproline II	rare	–	left	-78º	+149º
polyglycine II	rare	i···i+3	left	-80º	+150º

Table 3: *Geometric properties of strands within* β*-sheets*

Sheet Type	Frequency	H-bonding	Twist	Typical ϕ	Typical ψ
parallel	abundant	wide pair	right	-119º	+113º
antiparallel	abundant	close pair	right	-139º	+135º

The second major element of secondary structure is the β-**strand**. This is an elongated structure with backbone torsion angles not too far from the theoretical, fully extended conformation (φ = ψ = ω = +180º); as a result, both peptide groups and side chains are alternating between opposite orientations (**Table 3**). Obviously, such a linear structure does not allow for stabilization by local hydrogen bonds, as is the case with the α-helix. Instead, several β-strands can associate laterally to form a β-**sheet**, with a complete set of hydrogen bonds. These strands can be widely separated in the primary structure. Therefore, the β-sheet, as opposed to its strands, is not a secondary structure element in a strict sense[2] rather, it is an assembly of such elements and could hence be viewed as a super-secondary structure. It is important to note that the orientation of any two strands within a β-sheet can be either parallel or antiparallel, with the antiparallel form being slightly more stable due to a more favorable orientation of hydrogen bonding partners. When viewed along the direction of their strands, β-sheets usually display a right-handed twist. Both helices and strands are roughly linear structures and usually traverse the entire molecule (or domain) in a certain direction; therefore, they need to be linked by polypeptide segments located mostly on the protein surface. These loops or bends are generally rich in hydrophilic amino acids and often contain proline or glycine residues. For bends connecting β-strands, some characteristic conformations have been identified; the most common ones are classified as β-**turns** type I through III (and I' through III', with reversed φ-ψ signs). These β-turns are generally stabilized by a hydrogen bond between the carbonyl oxygen of residue n and the amide nitrogen of residue n+3. A different kind of bend with a particularly high curvature, characterized by an n··n+2 hydrogen bond, has been named γ-**turn**. While these bends or turns are constrained by a specific hydrogen bonding pattern, loops are larger segments with less regular structure, often displaying higher conformational freedom. An important class is the so-called ω-**loop**; it comprises up to 16 residues, with side chains typically sequestered in the loop center. In protein molecules, structures of this kind often confer important biological activities.

[2] In the IUPAC-IUB guidelines [2], secondary structure of a polypeptide segment is defined as "...the local spatial arrangement of its main-chain atoms without regard... to its relationship with other segments."

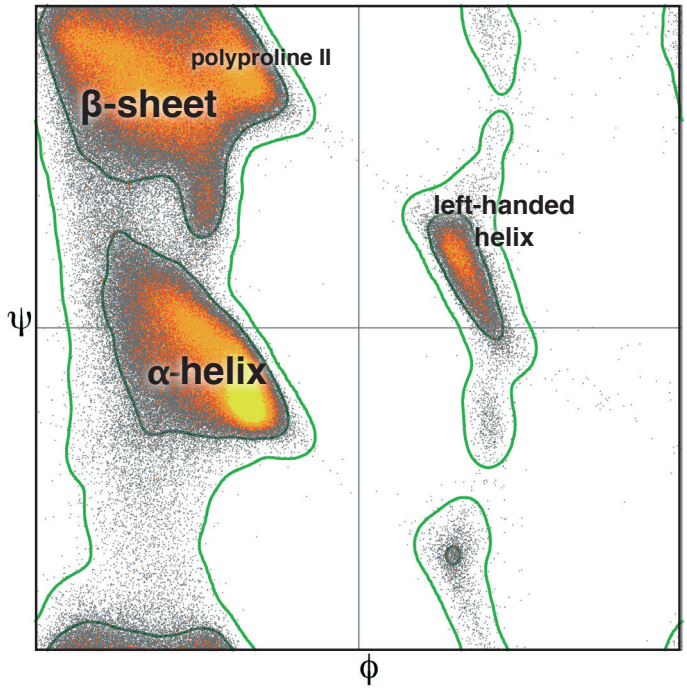


Figure 4: *The Ramachandran plot of torsion angles φ and ψ. Elements of secondary structure corresponding to the most favored regions are indicated. (Adapted from figure by Dcrjsr [CC BY 3.0] via Wikimedia Commons).*

As outlined above, secondary structure elements are usually stabilized by regular hydrogen bonding between peptide moieties and thus could, in principle, be independent of the amino acid sequence. However, the first experiments on in vitro protein folding clearly indicated that the final conformation of a polypeptide, including its secondary structure, is encoded exclusively in its sequence [1]. This is because side chains do interact with one another (and with backbone atoms), and the energetic contribution of such interactions depends on main chain geometry. Indeed, the preferences of different amino acids for certain areas in φ-ψ space can be used to predict the secondary structure of a polypeptide chain. The first algorithm of this kind has been developed by Chou and Fasman [3]. While this method essentially relies on the relative frequencies for different amino acids occurring in various types of secondary structure, the significance of the environment with respect to the local conformation has been recognized more recently. In modern algorithms for secondary structure prediction this is implemented e.g. by considering the conditional probability of an amino acid assuming a certain structure given that its neighbors assume the same structure (a case of Bayesian statistics).

It is interesting to note that the major secondary structure elements (α-helix and β-strand) have been predicted by Linus Pauling and coworkers as early as 1951, way before they were found experimentally by X-ray crystallography. Being a structural chemist working on three-dimensional structures of amino acids and short peptides, Pauling recognized the requirement for planar configuration of peptide groups; imposing this condition, he arrived at only two possible helical structures [4], one of which is now known as the α-helix. In the same year,

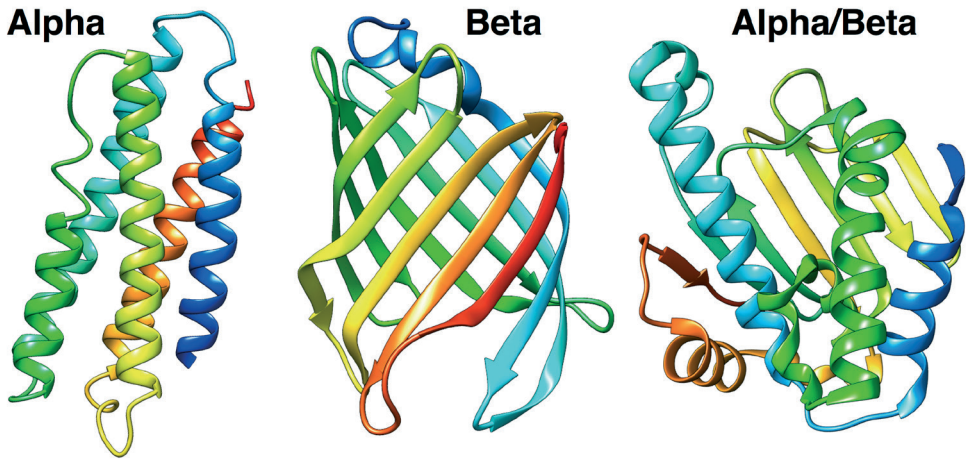


Figure 5: *The three principal classes of protein folds.*

the authors suggested the parallel and antiparallel β-sheet structures for extended polypeptide segments, with their distinctive hydrogen bonding patterns [5]. With the advent of protein X-ray crystallography, most aspects of their predictions turned out to be remarkably correct[3].

As a side note, these models implicitly defined the range of main chain torsion angles typically populated by folded polypeptides. A scatter plot of φ-ψ pairs realized in a given structure is called a **Ramachandran diagram** (Fig. 4). Notably, most values of rotation about the two bonds are forbidden for sterical reasons—of the two major allowed regions, one corresponds to the α-helix and one to the β-conformation.

Tertiary structure

The tertiary structure of a polypeptide chain describes the three-dimensional arrangement of its atoms in space. It can be thought of as the result of secondary structure elements packing together to form a more compact structure. In general, this can happen in virtually any way, although certain patterns are observed more frequently. Packing of two α-helices, for instance, preferentially occurs at an angle of either +20º or -50º, since these orientations favor indenting of side chains from the two segments. Similarly, β-sheets often combine to form sandwich structures at an angle of approx. 40º, while the mixed pair (α-helix with β-sheet) usually associates at about 30º. For secondary structure elements adjacent in the amino acid sequence, a number of characteristic assemblies have been recognized and termed **super-secondary structures**; examples include the β-hairpin (two antiparallel β-strands connected by a turn), the β-α-β structure (two parallel β-strands with their link containing a helical segment), and the helix-turn-helix motif (two helices arranged in any geometry, joined by a loop). Due to the architectural principles and boundary conditions outlined above, the size of compact protein

[3] There are only two notable exceptions: First, the figure printed in [4] shows a left-handed helix consisting of D-amino acids, which corresponds to the mirror image of the α-helix actually found in proteins. While the absolute configuration of natural amino acids was under investigation at that time, Pauling and colleagues obviously did not consider this ambiguity in their modeling approach. Second, their β-sheets [5] were completely flat, omitting the twist present in virtually all natural structures.

folds follows a relatively narrow distribution, mostly ranging from 40 to 100 (in rare cases exceeding 400) residues. Consequently, larger proteins usually contain several of these globular regions **(domains)**, which are thought to fold independently and are usually connected by more flexible linker sequences. Figure 5 illustrates the most common classification of domain structures, which is based on their secondary structure content.

Quaternary structure

In addition to solvent and small organic compounds, individual polypeptide chains often interact with other macromolecules. Such assemblies are referred to as quaternary structures. In general, they are stabilized by the same set of weak non-covalent forces as described above for the folding of single protein molecules: hydrophobic interactions, hydrogen bonds, van der Waals and electrostatic interactions. Protein-protein complexes can be conceptually categorized based on their life times and/or the mutual affinities of their constituents. In general, interactions can be permanent or transient; while permanent interactions are necessarily strong (such as many homo-oligomers which are virtually non-existent in monomeric state), transient ones can be intrinsically weak (i.e. affinity is constantly low) or strong (binding is triggered on demand, usually by reversible post-translational modifications). The latter case is extremely common in signal transduction networks, where protein-protein interactions need to be tightly regulated. A particularly impressive example is the mitotic cycle, which requires disassembly and reassembly of complex cellular organelles.

The concept of **surface complementarity** is crucial for understanding affinity and specificity in protein assemblies, for mainly two reasons: First, the elementary attractive interactions mentioned above are strongly dependent on the distance of the atoms involved; hydrogen bonds in particular also require a favorable orientation of participating orbitals. Second, significant affinity of binding partners requires several of those weak bonds to act in concert. The free energy contributions of individual contacts within a complex binding interface can, as a first approximation, be treated as additive; considering the relationship between standard free energy change and equilibrium constant ($\Delta G^0 = -RT \ln K$), the overall equilibrium constant is then obtained as the product of the constituent values. For practical purposes, the surface area which is desolvated upon formation of a macromolecular complex can be taken as a rough indicator for the strength of the interaction.

2.3 Post-translational modifications

Several of the amino acid side chains found in proteins can undergo covalent modification, usually catalyzed by specific enzymes. In most instances, these reactions occur after translation, i.e. in the folded polypeptide. For instance, secretory proteins as well as extracellular parts of membrane proteins often contain carbohydrate chains attached to serine/threonine (O-glycans) or asparagine residues (N-glycans); these can have a large effect on protein stability and functional activity. Another prominent example is **phosphorylation**, the reversible transfer of a phosphate moiety to side chains carrying hydroxyl or other functional groups. Indeed, phosphorylation is the most important regulatory reaction in living cells, governing the activities of countless enzymes and acting as a recognition signal in protein-protein interaction. Other modifications comprise hydrophobic molecules anchoring proteins to lipid bilayers, or chromophores conveying light sensitivity to their carriers.

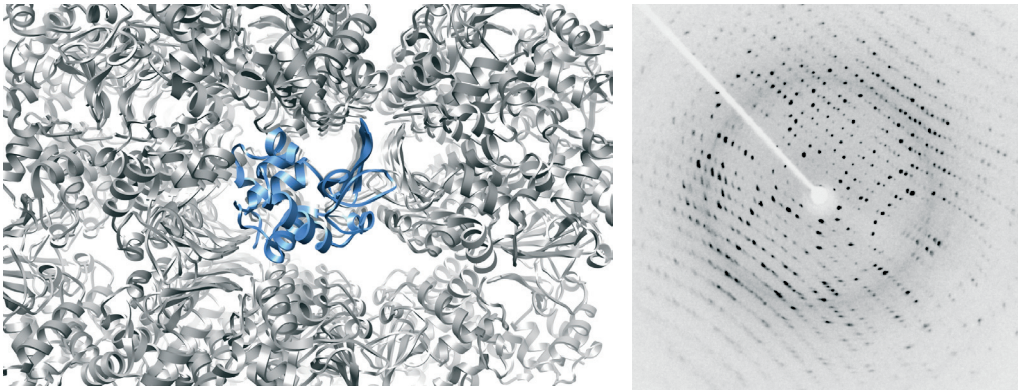


Figure 6 *(a) Showing the protein crystal packing of lysozyme in the spacegroup $P4_32_12$. (b) Showing the diffraction pattern obtained for a specific orientation of the lysozyme crystal.*

The sulfhydryl groups of cysteine side chains can be oxidized to form **disulfide bonds**, thus covalently linking residues not adjacent in the primary structure. Since the cytosolic milieu of cells is generally reducing, formation of disulfide bonds in vivo is restricted to proteins passing through the secretory pathway.

3 Structure Determination

Since the determination of the first protein structures by Perutz and Kendrew in 1960, the method of X-ray crystallography has become the main workhorse of structural biology to determine three-dimensional structures of biomolecules. The Protein Data Bank (PDB) [7] stores such atomic models of proteins and has more than 70,000 entries to date. X-ray crystallography has contributed more than 85% of all these PDB entries. This collection of structures represents an enormous amount of biologically highly relevant information and has to a large extent shaped our current view of the protein machinery. Atomic models help to understand intricate details of the molecular mechanisms that drive protein function. For the correct interpretation of these atomic models it is, however, important to understand how these models were obtained. While it is tempting to take the deposited structures as a perfect representation of the real macromolecules one should be aware of the fact that the models can have significant uncertainties or even severe errors. These uncertainties necessarily affect any conclusions drawn from these structures and in extreme cases can render them even meaningless.
The purpose of these lecture notes is to provide the user of the PDB with an idea of how the models were obtained and what their limitations are. To this end, the technique of X-ray crystallography is introduced, describing how to get from the measured data to an atomic model.

Since X-ray radiation scattered by a single protein is too weak to be detectable, a large number of (almost) identical copies of the same protein needs to be imaged at the same time. However, for the scattered light to add up in a meaningful way, the individual proteins need to be arranged in an ordered manner on a crystal lattice. In the following I will provide a compact description of how a protein structure can be determined by measuring the radiation

scattered from such a crystal. While these notes give only a very brief overview, the interested reader can find more details, e.g., in Refs [8,9].

3.1 Protein Crystallography

The first step in X-ray structure determination is to crystallize the protein that is to be studied. This is in many cases tedious and labor intensive as there are no rules or common procedures that will lead to success. Instead a large number of different parameters will have to be optimized to find conditions mostly by trial and error that eventually yield protein crystals. The parameters that are to be changed can be physical, chemical, or biochemical. Physical parameters include, e.g., temperature or pressure. Chemical parameters that affect crystallization are, e.g., salt concentrations, pH or the addition of specific ions or polymers like polyethylene glycol (PEG). In addition, it is often necessary to remove (supposedly flexible) parts of the protein by proteolysis.
Figure 6 (a) shows as an example how the protein lysozyme arranges in a crystal. Compared to small molecule crystals the packing in protein crystals is often relatively loose with only few crystal contacts. This leads in general to a quite high water content of about 50% on average (although this amount can vary greatly). The crystal is usually cooled to liquid nitrogen temperature (77K) to reduce radiation damage and thermally induced disorder and is then placed into an X-ray beam. The diffracted radiation is collected on film or on a CCD camera. The crystal is then rotated in small steps and for each orientation a diffraction image is recorded (see Figure 6 (b)).

If one assumes that atoms are spherically symmetric and that the electron density between atoms can be neglected then the scattered light wave of a single unit cell containing n atoms is described by the structure factor

$$F(\mathbf{S}) = \sum_{j=1}^{n} f_j \exp(2\pi i \mathbf{r}_j \cdot \mathbf{S}) \tag{1}$$

where $\mathbf{r}_j$ is the position of atom j and $\mathbf{S}$ is a vector perpendicular to the plane reflecting the incident beam, f_j is the atomic scattering factor and determines the scattering amplitude that is contributed by atom j. A more general formulation for the structure factor is given by

$$F(\mathbf{S}) = \int_{unitcell} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r} \cdot \mathbf{S})\, d\mathbf{r} \tag{2}$$

where $\rho(\mathbf{r})$ is the electron density distribution. From Eq. (2) it can be seen that the scattered light field is actually just the Fourier transform of the electron density.

The wave scattered by an entire crystal with $(n_1 \cdot n_2 \cdot n_3)$ unit cells is then the sum over all cells:

$$F_{\text{cryst}}(\mathbf{S}) = F(\mathbf{S}) \sum_{u=0}^{n_1} \exp(2\pi i u \mathbf{a} \cdot \mathbf{S}) \sum_{v=0}^{n_2} \exp(2\pi i v \mathbf{b} \cdot \mathbf{S}) \sum_{w=0}^{n_3} \exp(2\pi i w \mathbf{c} \cdot \mathbf{S}) \tag{3}$$

where **a**, **b**, and **c** are the basis vectors of the unit cell. For increasing crystal size the exponential terms in Eq. (3) become delta functions, i.e., $F_{cryst}(\mathbf{S})$ will only be non-zero if the following so called *Laue condition* is fulfilled

$$\mathbf{a}\cdot\mathbf{S}=h,\quad \mathbf{b}\cdot\mathbf{S}=k,\quad \mathbf{c}\cdot\mathbf{S}=l \tag{4}$$

where h, k, and l are integer numbers. That means only those light waves add up constructively that are scattered into certain directions, perpendicular to planes in the crystal which are defined by the integer triplet **h**=(*hkl*). This is the reason why the diffraction pattern shows individual peaks instead of a continuous distribution. The position of these peaks (the diffraction pattern) depends only on the arrangement of the proteins in the crystal, i.e. the spacegroup of the crystal, and not on the structure of the protein itself. The structure of the protein is encoded in the strength of the reflections, i.e. the amplitude of the structure factors.

Using the relation in Eq. (2) it turns out one can calculate the electron density from the measured structure factors (peaks in the diffraction pattern) by inverse Fourier transform:

$$\rho(\mathbf{r})=(1/V)\sum_{\mathbf{h}}F(\mathbf{h})\exp(-2\pi i\mathbf{r}\cdot\mathbf{h}) \tag{5}$$

By summing over **h**=(hkl) only those **S** are selected for which the Laue condition (Eq. (4)) is fulfilled.

In Figure 6 (b), the farther away the peaks are from the center (the larger the scattering angle) the smaller is the distance between the reflecting crystal planes. Those peaks therefore contain information on higher resolution details.

Resolution

As can also be seen from Figure 6 (b), the scattering intensity drops towards larger scattering angles (higher resolutions). The resolution is therefore defined as the largest scattering angle (or distance between reflective crystal planes) for which a complete set of reflection peaks can be detected.

Figure 7 shows electron density maps at resolutions between 1 and 4Å, to give an idea of what structural details are visible depending on the resolution. Hydrogens become visible only at resolution higher than 1Å and there are only few protein structures solved to a resolution that is significantly higher than 1Å. At resolutions beyond 3Å, which for X-ray crystallography can be considered already low resolution, individual atoms cannot be placed anymore with high confidence, even entire side-chains are not always visible. Figure 7 shows simulated electron density maps that are computed directly from the PDB and are therefore ideal. Electron density maps that are determined from experimental data will be less clear, especially at low resolution. Since $F(\mathbf{h})$ is a complex function, knowledge about both amplitude and phases is necessary to calculate the electron density. However, in the experiment only the intensity of the diffraction peaks can be measured. The intensities $I(\mathbf{h})$ are the squares of the structure factor amplitudes

$$I(\mathbf{h})\propto\left|F(\mathbf{h})\right|^2 \tag{6}$$

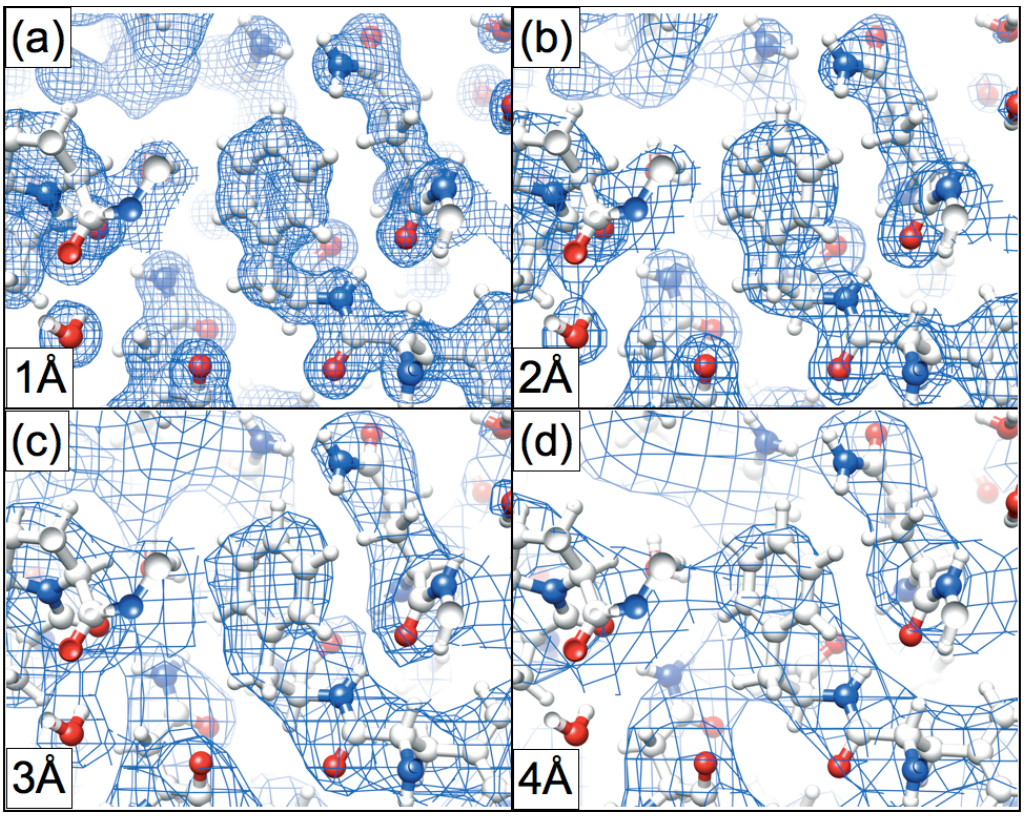


Figure 7: *Showing the level of detail visible in electron density maps at resolutions between 1Å and 4Å. Those maps are computed from synthetic data and are therefore ideal, whereas electron density maps derived from experimental will necessarily contain noise and additional artefacts not visible here.*

The phase information, however, is lost, which is refered to as the phase problem in crystallography.

3.2 Phasing

The phases of the structure factors can be determined experimentally by using basically one of the two techniques: 1) *isomorphous replacement*, where heavy atoms are typically soaked into the crystal and change the diffraction pattern from which phases can be reconstructed, and 2) *anomalous dispersion*: modern synchrotrons offer the possibility to use different wavelength which can exploit the anomalous scattering of some special elements. For example selenomethionine is an often used derivative of the regular amino acid methionine where the sulfur is replaced by the anomalous scatterer selenium. From the differences in the diffraction pattern obtained with different wavelengths phases can be calculated. These experimental phasing techniques, however, require additional, often very time consuming experiments. In the case where a similar structure (or at least a fragment of it) is already known, this structure can be taken as an approximate solution from which approximate phases

can be obtained using a technique called *molecular replacement*. Because of the large number of known protein structures this method has become very popular.

3.3 Model Building

Once initial phases could be obtained, an electron density map can be calculated using Eq. (5). Depending on the quality of the initial phases the electron density shows a more or less clear picture of the protein. All modeling typically starts with identifying the trace of the backbone (main-chain). Based on this trace the side-chains are then placed. As the amino acid sequence of the protein is usually known, the task is to assign the sequence to the backbone. This assignment is guided by the electron density. For lower resolutions (>3Å) it can in fact happen to introduce register shifts in certain regions where side-chain densities are not clearly distinguishable. Automatic modeling tools (e.g. ARP/wARP [16], RESOLVE [17], TEXTAL [18]) exist to facilitate the tedious work of building an initial model. The success of such modeling tools, however, heavily depends on the resolution and the quality of the initial phases. For example ARP/wARP can build on average more than 90% of the amino acids correctly if the resolution is better than 2Å; for a resolution of 3Å this value drops to 74%. Building the model always needs considerable additional manual interaction. In particular, loop regions require manual adjustments as they tend to be more flexible and as a result their density is usually not as clearly resolved. Modeling loops is still a considerable challenge for automatic model building programs (as well as for humans).

Once an initial model has been built into the density map, this model is further refined by fitting the model (i.e. the atomic coordinates) against the measured structure factor amplitudes. In the next step new phases are computed from the refined model, which yields a (hopefully) improved electron density map, where new features might become visible which can be used to build missing parts of the model or to further optimize the model.

The general workflow is an iteration over model building, structure refinement, and computation of an updated electron density map. The iteration is continued until the model converges and cannot be significantly improved anymore by further refinement.

3.4 Refinement

The straight-forward approach to refinement is to minimize the deviation of the observed data $F_{\mathrm{obs}}(\mathbf{h})$ from the data calculated from the model $F_{\mathrm{calc}}(\mathbf{h})$. If no experimental phases are available, the amplitudes of the structure factors of the model need to be in agreement with the observed amplitudes, i.e. the function

$$L = \sum_{\mathbf{h}} \left\| F_{obs}(\mathbf{h}) \right| - \left| F_{calc}(\mathbf{h}) \right\|^2 \tag{7}$$

is minimized with respect to the atomic coordinates. This least-squares type of refinement is however not used anymore since it has been shown that maximum likelihood approaches [10,11] are superior as they can more rigorously take into account error probability distributions and partial models and eventually show less artifacts than the least-squares method. The agreement of the model with the data is often measured with the R-factor

$$R = \frac{\sum_{\mathbf{h}} \left| |F_{obs}(\mathbf{h})| - |F_{calc}(\mathbf{h})| \right|}{\sum_{\mathbf{h}} |F_{obs}(\mathbf{h})|} \tag{8}$$

which is a value between 0 and 1. The refinement seeks to lower this R-value. The simplest optimization strategy is to perform a steepest descent minimization of the target function L.

A computationally more expensive approach is simulated annealing molecular dynamics simulation [15], which has been shown to significantly increase the radius of convergence by helping to escape local minimia.

The thermal motions of atoms reduce the intensity of the observed reflections, which has to be accounted for in the refinement. Oftentimes these motions are approximated to be isotropic, such that the structure factor is written

$$F(\mathbf{S}) = \sum_{j=1}^{n} f_j^0 \exp(-B_j |\mathbf{S}|^2) \exp(2\pi i \mathbf{r}_j \cdot \mathbf{S}) \tag{9}$$

where B_j is the atomic B-factor of atom j. It can be shown that the B-factor is related to the mean-square displacement of the vibrating atom, $\overline{u^2}$, by

$$B = 8\pi^2 \overline{u^2} . \tag{10}$$

The atomic B-factors are usually also refined along with the atomic coordinates. This increases the number of parameters that are to be determined.

Parameter to Observation Ratio

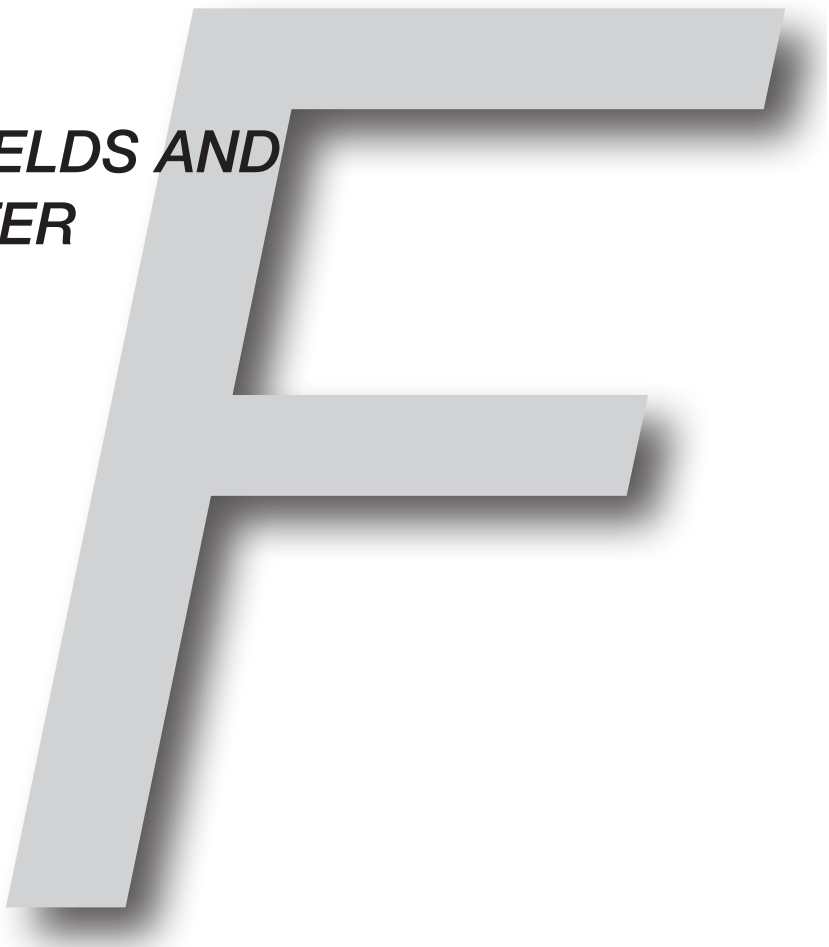
The number of observations in an crystallographic measurement (number of independent reflections) depends strongly on the resolution. The number of parameters (atomic coordinates and isotropic atomic B-factors) that are to be determined is typically large and exceeds the number of observations in an average case already at a relatively high resolution of about 2.5Å. There is however additional information that can be used to make the problem tractable. The knowledge of the sequence is, e.g., such an important additional restraint. Furthermore, the fact that proteins are composed of individual atoms (atomicity) provides a very strong restraint which allows to accurately determine a structure even when no phase information is available.

However, for resolutions that are significantly lower than the typical distance between atoms the effect of the atomicity vanishes. In this case one uses additional restraints on the bond lengths and angles to impose the atomicity. Those restraints can be applied without adding a strong bias since the local geometry of the bonds between specific atom types are to a good approximation the same in all proteins.

References

[1]C.B. Anfinsen, R.R. Redfield, W.L. Choate, J. Page, and W.R. Carroll, J. Biol. Chem. 207, 201 (1954).
[2]IUPAC-IUB Commission on Biochemical Nomenclature, J. Mol. Biol. 52, 1 (1969).
[3]P.Y. Chou and G.D. Fasman, Biochemistry, 13, 222 (1974).
[4]L. Pauling, R.B. Corey, and H.R. Branson, Proc.Natl.Acad.Sci USA 37, 205 (1951).
[5]L. Pauling and R.B. Corey, Proc.Natl.Acad.Sci. USA 37, 251 (1951).
[6]J. Kraut, Ann. Rev. Biochem. 46, 331 (1977).
[7]Protein Data Bank (http://www.rcsb.org)
[8]International Tables of Crystallography (http://it.iucr.org)
[9]J. Drenth, Principles of Protein X-ray crystallography (Springer, New York, 1994)
[10]G. Bricogne, Meth. Enzymol. 276, 361 (1997)
[11]N.S. Pannu and R.J. Read, Acta Cryst. A52, 659 (1996)
[12]S. Hovmöller, T. Zhou, and T. Ohlson, Acta Cryst. D58, 768 (2002)
[13]A.T. Brunger, Nature 355, 472 (1992)
[14]G.J. Kleywegt, M.R. Harris, J.Y. Zou, T.C. Taylor, A. Wählby, and T.A. Jones, Acta Cryst. D60, 2240 (2004)
[15]A.T. Brunger, J. Kuriyan, and M. Karplus, Science 235, 458 (1987)
[16]G. Langer, S.X. Cohen, V.S. Lamzin, and A. Perrakis, Nature Protocols **3**, 1171-1179 (2008)
[17]T.C. Terwilliger, Methods Enzymol. **374**, 22-37 (2003)
[18]T.R. Ioerger and J.C. Sacchettini, Methods Enzymol. **374**, 244-270 (2003)

EXTERNAL FIELDS AND ACTIVE MATTER



F 1 Structure and Flow

M. P. Lettinga
Institute of Complex Systems / Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Flow instabilities **4**

3 Flow profiling **6**
3.1 Photon Correlation Spectroscopy in flow 8
3.2 Particle Imaging Velocimetry 8
3.3 Ultra-sound Velocimetry 11
3.4 NMR 11
3.5 Optical coherence tomography velocimetry 12

4 Probing the structure **13**
4.1 Optical rheometry: polarization 14
4.2 Optical rheometry: light scattering 16
4.3 *In situ* Scattering: Neutrons 19
4.4 *In situ* Scattering: X-rays 20
4.5 *In situ* Microscopy 25

5 Conclusions **25**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Flow is one of the most prominent aspects of nature, albeit the flow in the oceans or the blood flow through the smallest of our veins and arteries. These two examples show the huge difference in length scales as well as the very different aspects of flow. Turbulence often plays an important role in the ocean, which is due to the inertia of the moving fluid. Of course turbulence also occurs at small length scales, for example in case of arteriosclerosis. Another difference is that the ocean consists of water, which is a simple Newtonian fluid, while blood is a complex fluid containing high concentrations of complex particles. Such fluids can display a very complex flow behavior where for example the viscosity strongly depends on the rate at which the fluid is deformed. As a consequence there are very diverging fields of research.

In this chapter we will focus on the flow behavior of complex fluids, as it is most relevant for functional soft matter. Throughout this book you will find many examples of complex fluids, as many subjects deal with some kind of particles dispersed in a fluid, which is mostly water. Whilst in most cases the focus is on the structure and dynamics of the particles, the question we ask in this chapter is how and why the structure and dynamics of the particles affects the response of the complex fluids it constitutes to deformation. This is in principle the main question in the field of rheology, which has already been introduced in chapter C6. In classical rheology the main tools to characterize the sample are (time-resolved) mechanical measurements of the stress and deformation. Although measurements can be done with incredible precision and relatively detailed structural information can be obtained from these experiments, there are a few limitations to a pure rheological approach. First, the structural information is obtained indirectly, so that any statement about the structure would need to be tested by techniques that directly probe these structures. Second, complex fluids can have complex flow behavior. This means that there is no simple linear flow profile between the moving parts of the rheometer, which hampers the interpretation of the results. To fully understand the flow of complex fluids it is therefore a prerequisite to have information on three different length scales: The macroscopic response needs to be known using rheological tools that may require milliliters of sample to measure sufficient torques imposed on one plane by the opposite moving plane, transmitted by the fluid; the flow profile of the sheared fluid needs to be probed, which is typical on the length scale of a millimeter; the structure of the dispersed particles needs to be probed on a micrometer to nanometer length scale. In the ideal case this structure is probed locally along with the local velocity. Thus, in an ideal rheological experiment the macroscopic mechanical response of the fluid is measured simultaneously with the microscopic structure and flow profile.

The goal of this chapter is to give a feeling of the experimental difficulties that need to be solved in order to perform the ideal experiment. This is done by means of a few benchmark complex fluids that display a complex flow behavior due to shear induced structural changes in the fluid and relates to the rheology introduced in chapter C6. We will start by introducing typical flow instabilities that may occur in a sheared sample and the relation with their rheological response. We then introduce several techniques that are used to measure the flow profiles and show examples. We finish by introducing techniques to probe the structure. We elucidating the pros and cons of the techniques depending on the fluid that is probed.

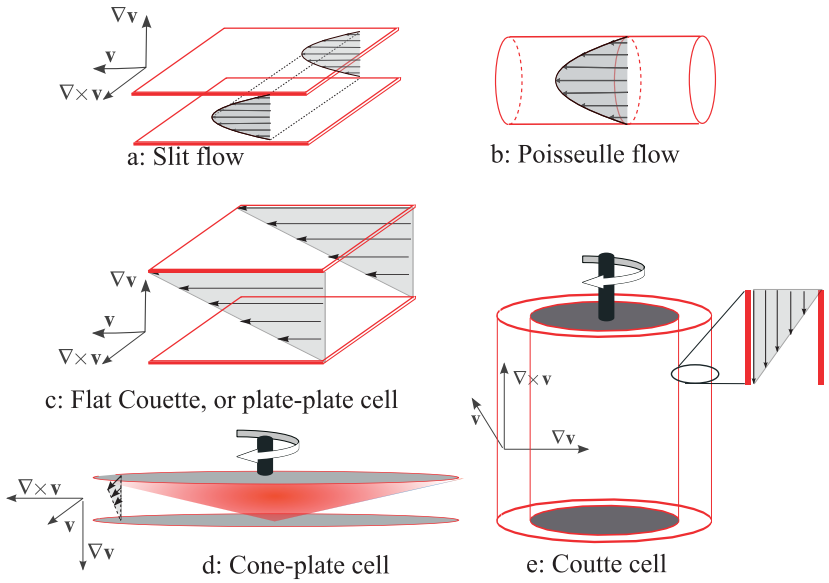


Fig. 1: *Standard geometries used in rheological experiments.*

2 Flow instabilities

For simple "Newtonian" fluids the stress is proportional to the applied shear rate, while for "non-Newtonian" fluids this is not the case. "Non-Newtonian" response sets in when the applied shear rate is faster than the relaxation dynamics of the system. If this is the case, then the system will be perturbed by the shear rate and change its properties. The direct consequence will be that the viscosity of the sample will change with the shear rate that is applied. This will be apparent when taking a flow curve where the stress, and hence the viscosity, is measured as a function of shear rate. The simplest flow curve for Newtonian fluids is a straight line where the viscosity does not change with shear rate or the stress increases linearly with shear rate. If the flow is not too fast and inertia of the fluid can be neglected, as we discussed in the introduction, then the flow profile of a Newtonian fluid is linear and can be thought of as sliding layers of fluids that have incrementing velocity. According to the definitions used in this field, the flow direction $\mathbf{v}$ is in the x-direction and the gradient in the flow, the gradient direction $\nabla\mathbf{v}$, is in the y- direction. The z-direction, the vorticity direction $\nabla \times \mathbf{v}$, is often referred to as the neutral direction since there is no convection nor distortion in that direction. In Fig. 1 we show the flow profile for different geometries that are used in rheology.

Two examples of extreme non-Newtonian flow behavior are given on the right of Fig. 2. In the top graph the stress jumps from σ_1 to σ_2 at a critical shear rate of $\dot{\gamma}^*$, and hence at this shear rate the viscosity jumps to a higher value. Increase of the viscosity with increasing shear rate is called shear thickening. In the bottom graph a region of shear rates can be identified between which the stress does not change, or where the shear rate jumps from γ_1 to γ_2 at a critical stress of σ^*. Hence, at this stress the viscosity jumps to a lower value. Decreasing viscosity with increasing shear rate is called shear thinning. Strong shear thinning is a very desirable feature for industrial applications. The flow behavior of shampoo, for example, is mainly engineered by using surfactant wormlike micelles (WLMs): at low shear it behaves like a gel, at high shear

it flows very easily. The same features are also very useful in, e.g. the oil industry [2]. Shear thickening is used for dampening materials like body armor. When a bullet enters a jacket with shear thickening material at high speed then the material stiffens and the bullet is slowed down [3].

When the flow behavior is extremely non-Newtonian then the stress behavior can cause mechanical instabilities in the shear cell. This means that instead of a smooth linear flow profile between the moving wall and the standing wall, as is required for Newtonian flow, the flow will be structured in either the vorticity direction, the gradient direction or both. Flow instabilities can be predicted on the base of constitutive models that describe shear thickening and thinning behavior. Constitutive models that describe extreme shear thickening assume a functional form as depicted in Fig. 2a, where the dashed line represents meta- or unstable states. A stable state, assuming laminar flow, would be achieved when alternating bands are formed with different viscosity, stacked along the vorticity direction [4]. The bands have a different stress as they are subjected to the same shear rate, as indicated by the dotted lines in the right panel of Fig. 2a. This "classical" picture of vorticity banding is shown schematically on the left of Fig. 2a. The question is, however, what mechanism drives the formation of these bands and whether the flow within the bands is laminar with equal shear rates. Constitutive models that describe extreme shear thinning again assume a functional form as given in Fig. 2b, where again the dashed line represents meta- or unstable states: the part where the stress decreases with increasing shear rate renders the system mechanically unstable. "Unstable" means in this case that small deviations from the linear velocity profile will grow. From a theoretical point of view this is very similar to an equilibrium gas-liquid phase separation [5, 6]. In the 'classical' picture of gradient band formation, at the end of the shear driven phase separation, the system splits up into a region of a low viscosity and high shear rate close to the moving wall and a region with a high viscosity and low shear rate close to the static wall such that the stress in both bands is constant, see the left of Fig. 2b.

Gradient banding in shear flow is far better documented [1] and understood [7, 8] than vorticity banding. In principle its occurrence is independent of the structural changes in the fluid underlying the shear thinning behavior, as long as the shear thinning is substantial. Thus gradient banding has been observed for systems ranging from hard sphere colloidal crystals [9] and soft colloidal glasses [10] to associative polymer networks [11], entangled polymer solutions [12, 13] and DNA solutions [14, 15]. Shear thinning behavior is far more pronounced for dispersions of living polymers, which consist of monomers that continuously break and recombine. This class of polymers has two mechanisms to release stress: reptation, as discussed in chapter B1, and the break up and recombination of the polymers. For the case where the average breaking time is much faster than the reptation time it can be shown that this results in a unique relaxation mechanism and strong shear thinning [16, 17]. Surfactant wormlike micelles (WLMs) are self-assembled particles that display strong shear thinning, which is easily tunable [18, 19]. Therefore WLMs are the 'working horse' for experimentalists in the field of flow instabilities. Throughout the chapter we will show many examples of measurements on WLMs. The persistence length of WLMs is much longer then the thickness of the WLMS and therefore these systems display strong alignment effects when subjected to shear flow.

Vorticity banding is less common than gradient banding. A detailed recent discussion of elastic instabilities that drive the formation of vorticity bands can be found in Ref. [20]. The classical picture of vorticity banding due to shear thickening is more the exception than the rule, although again for wormlike micelles the predicted behavior has been found [21]. Another, more common, mechanism is that normal stresses are generated through the non-linear elastic

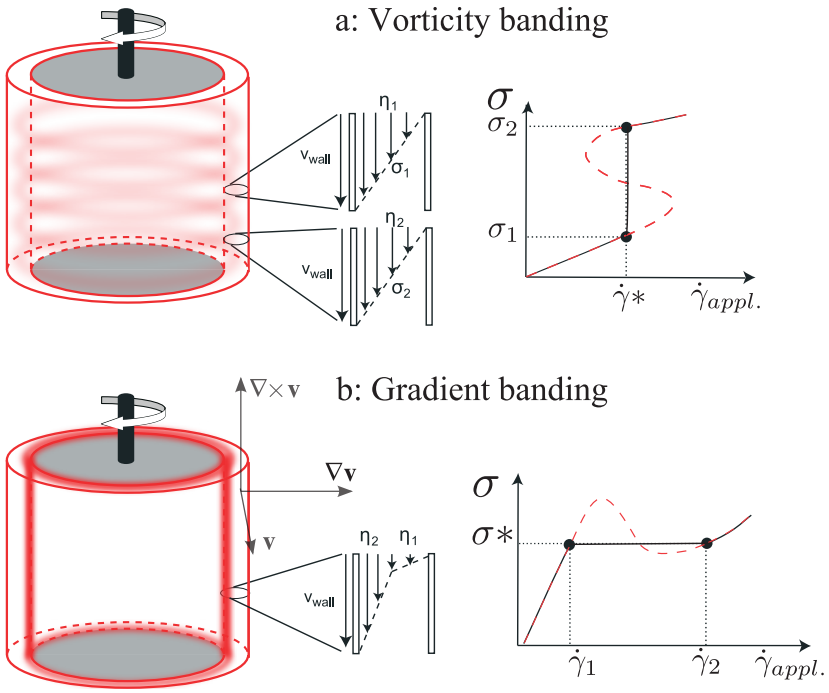


Fig. 2: *Two types of shear banding: in the vorticity direction (a) and in the gradient direction (b). The flow profiles are indicated in the middle by the different lengths of the arrows. To the right flow curves underlying the shear band formation at the right.*

deformation of the dispersed mesoscopic entities, similar to the Weissenberg effect in polymeric systems [22]. The role of the polymer chains in the classic Weissenberg effect is now played by large scale networks, from for example carbon nanotube dispersion [23] or clay gels[24], or by inhomogeneities formed during the initial stages of phase separation as found for polymer-protein mixtures [25] and colloidal rods [26]. These structures will be non-uniformly stretched due to the curved streamlines that are present in most rheological devices, thus generating hoop-stresses that give rise to elastic normal forces [27]. Another mechanism for the formation of vorticity bands shear bands actually starts with the formation of of gradient bands. There is experimental evidence [28, 29] and theoretical justification [30] that vorticity bands are formed due to instabilities in the interface between two gradient bands. All of these scenarios lead to rolling motion within the bands, superimposed on the laminar flow. This rolling motion is absent in the classical picture, depicted in Fig. 2a. Note also that these are pure elastic instabilities and that inertia does not play a role, as is the case for well known Taylor instability [31].
Clearly the field of flow instabilities is very rich and requires an extensive experimental as well as theoretical toolbox to make progress in this field. The examples given above illustrate why this field is not only of practical but also of fundamental interest. It can be that the thermodynamic phase behavior of the dispersion is coupled to its hydrodynamic response, such that only small perturbations can lead to big effects. On the other hand also the constituting particles can change their features due to the applied shear inducing flow instabilities. We will now first show how flow instabilities can be identified in section 3 and then show in section 4 how the underlying microstructure can be measured *in situ*, i.e. while shearing the system.

3 Flow profiling

Ideally the flow is measured in all directions with the highest possible time and spatial resolution. Indeed, in recent years this ideal is approached by techniques such as ultra-sound velocimetry and particle imaging velocimetry. Both of these techniques, however, make use of tracer particles that might distort the rheological response. Other techniques use the inherent contrast of the sample, such as heterodyne light scattering, but such optical techniques are restricted to optical transparent samples. We will now treat the different techniques mainly on the hand of studies of shear banding wormlike micelles, as for these systems shear banding behavior is very reproducible.

3.1 Photon Correlation Spectroscopy in flow

In a homodyne dynamic light scattering, coherent light scatters from a fluctuating sample, giving rise to intensity fluctuations due to a changing speckle pattern. These intensity fluctuations can be analyzed by means of the intensity autocorrelation function $g_2(q, \tau)$, where τ is the lag time between different events and q is the scattering wave vector, see chapter C1. In the case of a fluid in laminar flow, the autocorrelation functions are determined by three factors: a) The diffusive motion of the scatterers due to thermal fluctuations in the sample, characterized by the diffusion coefficient D; b) the transit time of the particles in the scattering volume, depending on the size of the scattering volume h and the average transit speed of the particles v_t; c) Doppler shifts arising due to different average flow velocities of scatterers, characterized by the shear relaxation rate Γ_s. Thus, in principle the local shear rate and velocity can be point wise measured [34, 35]. This does, however, require knowledge of the exact shape of the scattering

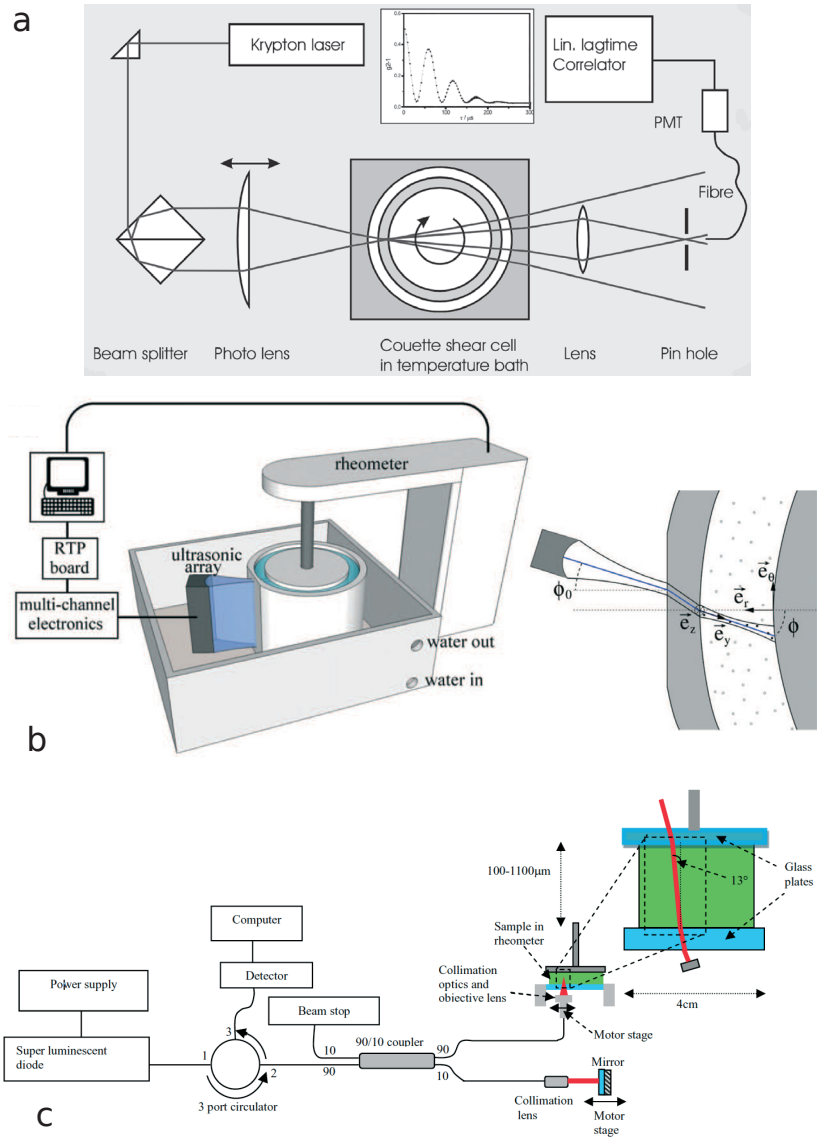


Fig. 3: *Schemes of the instruments used for measuring the velocity profiles. (a) Heterodyne light scattering using two crossing incident laser beams. The beam intersection length in the radial direction is* $\approx 100\mu m$. *Image taken from Ref. [32]. (b) Ultrasound Velocimetry using 2D transducers with which 2D profiles can be taken at very high rates, but with the same spatial resolution as the heterodyne light scattering. Image taken from Ref. [33]. (c) A schematic diagram of the optical coherence tomography shear rheometer with a plate-plate cell. Image taken from Ref. [47]*

volume and the intensity profile of the incident beam. Moreover, the interpretation becomes ambiguous in case of a complex shear banding flow.
Laser Doppler velocimetry uses Heterodyne Dynamic Light Scattering to obtain a local measurement of only the velocity. Here the scattered light is mixed with a reference beam so that $g_2(q,\tau)$ reflects the interferences between the Doppler-shifted light that has crossed the sheared sample and the reference light. In Fig. 3a we display a set-up where actually two lasers beams are crossed in the gap of a Couette cell where they produce an interference pattern over a depth of approximately 100 μm. When a scatterer passes through this volume, it will scatter light from each of the two beams with a different Doppler shift. The average velocity inside the scattering volume is directly obtained from the frequency of the resulting oscillation in $g_2(q,\tau)$, as shown in the little graph in Fig. 3a. The flow profile is obtained by scanning the scattering volume through the gap. The time to take a flow profile in e.g. a Couette cell is about half a minute, as it takes at least a couple of seconds to obtain a accurate autocorrelation function and at least 10 points need to be taken to produce a flow curve over a gap of 1 mm. The rate of data acquisition can be enhanced by adding strongly scattering tracer particles. This might, however, affect the system, as we mentioned earlier. Another disadvantage of the technique is that it requires optically transparent samples, at least over the gap. Still, it is a relatively popular technique because it is relative easy to set up. Salmon et al [36] demonstrated the use of this technique. Here it was clearly demonstrated for the first time that strongly shear thinning WLMs indeed split up in two bands with shear rates of $\dot{\gamma}_1$ and $\dot{\gamma}_2$, see Fig. 2b, following a lever rule predicted by theory [16, 37, 6], similar to the lever rule for gas-liquid phase separation.

3.2 Particle Imaging Velocimetry

The lever rule for the same system of WLMs was confirmed only a couple of years later, using Particle Image Velocimetry (PIV) [43, 39]. In PIV the local velocity is determined by measuring the displacement of a particle between two sequential frames. This is often done by using a sheet of laser light shining through the gap of the shear cell, in combination with a fast camera oriented in the direction of the normal of this light sheet. A high time resolution in the order of milliseconds or higher can be achieved, depending on the camera, and a spatial resolution of less than 100 μm, depending on the optics and tracer used. Using a laser sheet that is oriented in the plane of the flow, the formation of shear bands could be followed from the moment the shear rate is switched on [39]. If the temporal resolution is high enough, the propagation of flow through the gap of the shear cell can be resolved, see Fig. 4c. It was also shown that shear bands form in the middle of the gap and do not nucleate from the side, as was often assumed.
The 2-D motion can be obtained when orienting the laser sheet perpendicular to the flow plane, which is a prerequisite for the understanding of vorticity banding. It was shown in this way that the wormlike micelle system can undergo a transition to a complex rolling motion that is connected to structure formation in the vorticity direction [42]. The resulting profile is shown in Fig. 5b, together with a scattering image. Vorticity bands that are not related to gradient banding, such as observed when shearing the isotropic-nematic phase coexistence of rod-like virus also display complex rolling motion, which shows that the schematic picture of vorticity banding as given in Fig. 2 is at least idealized. PIV can be extended to three dimensions,

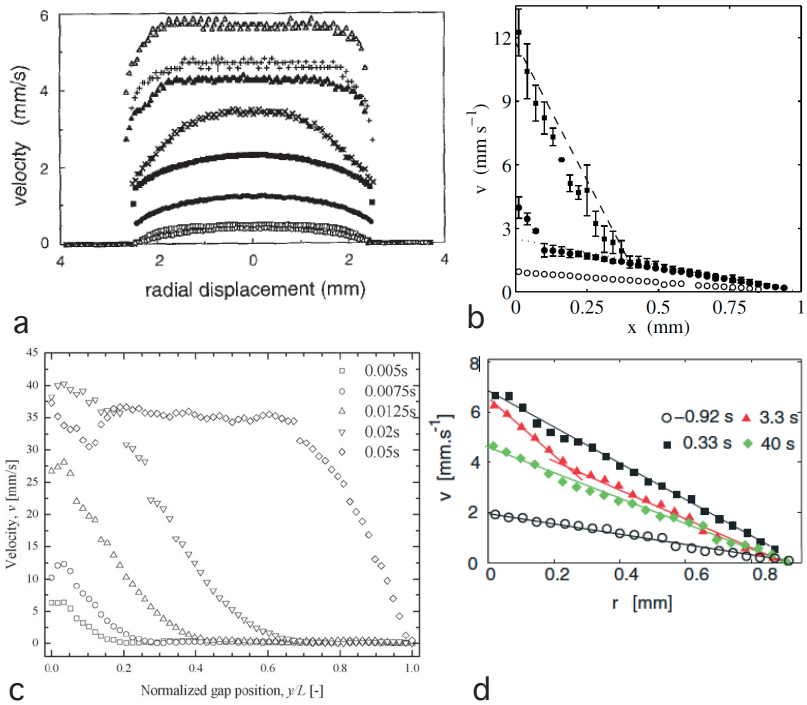


Fig. 4: *1D flow profiles of exactly the same surfactant WLMs taken by different techniques: (a) Poiseuille flow in a tube, see Fig. 1b, using NMR. Image taken from Ref. [38]; (b) Velocity profiles recorded using dynamic light scattering in heterodyne mode at a shear rate of* $\dot{\gamma} = 1\ s^{-1}$ *(○), showing a linear flow profile, and* $\dot{\gamma} = 5\ s^{-1}$ *(●) and* $\dot{\gamma} = 12\ s^{-1}$ *(■), showing shear banding. Image taken from Ref. [36]. (c) Transient velocity profiles taken by Particle Image Velocimetry from a startup experiment to an imposed shear rate of* $\dot{\gamma} = 8\ s^{-1}$*, showing the propagation of the flow into the gap of the cell. Image taken from Ref. [39]. (d) Transient velocity profile taken by Ultra-Sound Velocimetry , showing how initially bands form after applying a shear rate of* $\dot{\gamma} = 8\ s^{-1}$*. Here the bands disappear due to slip, which can be seen from the fact that the velocity at the moving wall is reduced. Image taken from Ref. [40]. The flow profilies in (b-d) are all taken in Couette geometry, see Fig. 1e, and have the moving wall at the left and the standing wall at the right.*

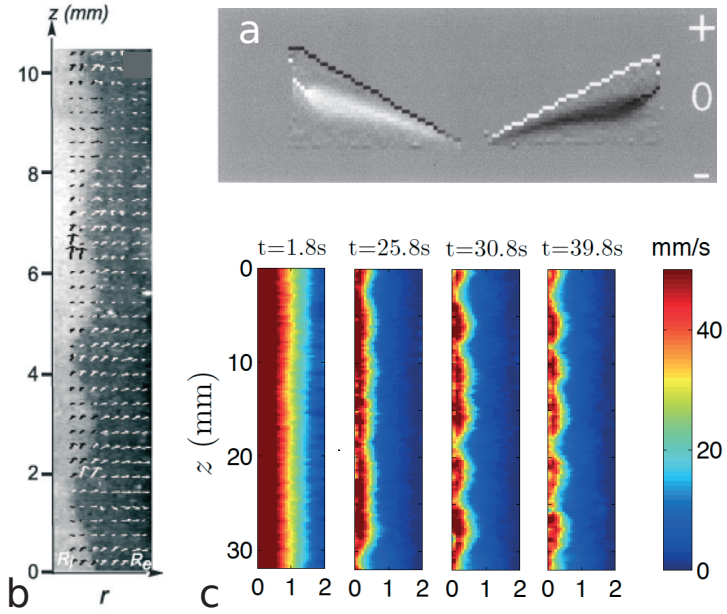


Fig. 5: *.2-D flow profiles on surfactant WLMs by different techniques. (a) Using NMR in combination with a cone-plate cell with a diameter of* $28\ mm$ *and a cone angle of 4 at a shear rate of* $\dot{\gamma} = 16\ s^{-1}$*. The grey scale indicates the shear rate in arbitrary units. Note the opposite sign shear for the receding and advancing segments of fluid on opposite sides of the gap. Image taken from Ref. [41]. (b) Stationary banding structure and corresponding velocity field computed from the PIV algorithm in combination with a Couette cell at a shear rate of* $\dot{\gamma} = 40\ s^{-1}$*. Image taken from Ref. [42]. (c) Formation of shear bands after start-up of shear in shear in a solution of WLMs seeded with hollow glass spheres at 1 wt. % and sheared at* $\dot{\gamma} = 46\ s^{-1}$*, using USV. Velocity maps at different times t indicated on the top row. Each map corresponds to an average over 100 pulses sent every 1 ms. Image taken from Ref. [33].*

using laser sheets with different orientations as well as several cameras. In principle confocal microscopy can also be used, where particles are images in 3D. In section 4.5 we will come back to this technique, which is mainly used to probe structures.
There are two main draw backs for the use of PIV. First, PIV is mostly imaging tracer particles and thus does not give any structural information. For this reason PIV was combined with small angle light scattering and birefringence imaging in Refs. [43, 39]. Second, the sample needs to be optical transparent, which strongly limits the number of systems that can be studied.

3.3 Ultra-sound Velocimetry

Ultra-sound Velocimetry (USV) is in many ways similar to PIV, except that it can be used for opaque samples as it employs sound to probe the flow. As with PIV, it uses a repetition of short sound waves to locate the particles, using the delay in the echo, and determine their displacement, using the shift in the echo between two sequential pulses. As with PIV it is routinely used in combination with a Couette geometry, see Fig. 5b [44]. USV experiments on the wormlike micelles yield very similar results as the PIV measurements. As USV can be employed for opaque systems, it was used to study the effect of wall slip by comparing the flow behavior in smooth (transparent) cells and sanded (opaque) cells, which obviously cannot be done by PIV. Fig. 4d shows the response after start up of flow for a smooth Couette cell. As the left boarder of the plot corresponds to the velocity of the sample at the wall, it can be directly read from the value at that point if the systems undergoes slip or not. It can be observed that initially this values is high and that bands form, but that later this value drops and the bands disappear. Thus, slip is in competition with the formation of shear bands [40].
These measurements were performed with one single transducer so that a 1-D profile can be obtained. Recently arrays of transducers are being used so that complex flow as described above in section 3.2 for the gradient and vorticity banding WLMs can be straightforwardly obtained, as can be seen in Fig. 5c [33]. With this recent development, USV became a very powerful technique to measure flow, despite of the draw backs that no structural information is obtained, while trace particles need to be used in the experiments.

3.4 NMR

It could be stated that potentially Nuclear Magnetic Resonance (NMR) Microscopy is the most complete technique as it can be used to obtain information about flow as well as structure of the complex fluid. Information about both nuclear spin positions and nuclear spin velocities is acquired via the precession of nuclear spins in a magnetic field. Because the nuclear spin precession frequency is dependent on the strength of the magnetic field in which it is placed, it is possible to determine the location of these spins by employing spatially-dependent magnetic fields. Magnetic field gradients can also be used to reveal molecular translational motion, simply by carefully measuring precessional phase changes. Proton NMR is the favoured mode of magnetic resonance imaging, because the hydrogen nucleus provides the most sensitive spin for nuclear magnetic resonance, . The predominant signal arises from water molecules which thereby provide a marker for the velocity of elements of the viscous fluid. NMR has played a very important role in the field of complex fluids, as the existence of shear banding for WLMs has first been shown by Callaghan et al. in a pipe-flow geometry [38], see Fig. 4a and later in cone-plate geometry [41], see Fig. 5a.

NMR has the additional advantage that detailed information can be obtained on the smallest relevant length scale. For example, the diffusion rate of the lipids that constitute the WLMs can be obtained, from which the degree of branching of these systems can be inferred [45]. The number of these mobile branches greatly influences the rheology of the system as it supplies an alternative relaxation route. Moreover, the local concentration and orientation can in principle be obtained [46]. Note that all of these features cannot be obtained in one and the same measurement. Different techniques like deuteron-NMR and pulsed gradient stimulated echo NMR are applied to obtain different features.
Summarizing, NMR does not need any tracer particles, supplies information on the flow as well as on the structure and can be used for opaque samples. Thus, it is a very powerful technique. The only drawbacks are that it is relatively slow and rather expensive. As the technical aspects continuously improve, given the broad range of applications, it is anticipated that both drawbacks will gradually disappear over time. Another drawback is that it is difficult to measure the rheology together with the NMR, because the shear cell needs to be submerged in a big magnet, which requires that the rotating shaft is very long.

3.5 Optical coherence tomography velocimetry

In Optical Coherent Tomography (OCT) velocimetry infrared light is used to probe the structure of opaque materials. Infrared light has the advantage that it scatters much less than visible light and can therefore penetrate opaque samples. It is frequently applied in medicine for imaging. Very recently it has been introduced in rheology to measure flow profiles [47, 48, 49].
In OCT the interference between infrared light that is scattered from an objected in the sample and reflected from a reference mirror is measured. Interference will occur when the path length difference between a particular depth within the sample and the reference beam is within the coherence length, which is typically less then 10 μm. This is similar to the echo in USV, except that the resolution is much better with a coherence length of less than $10\ \mu m$ and a beam-waist of about $20\ \mu m$. Different depths within the sample can be probed when the reference mirror position is varied along the beam path. This is similar to the delay in the echo for USV. If the sample is moving under the effect of shear imposed by the rheometer, the Doppler shift of the quasi-elastically scattered light will cause the interference fringes corresponding to a particular depth within the sample to move with a frequency that is proportional to the velocity of the sample. This is similar to the shift in the echo for USV. Thus simply by moving the reference mirror, a flow profile can be taken in about 10 seconds. A typical set-up is shown in Fig. 3c.
OCT is especially suited for opaque samples, similar to USV and NMR, because of the use of infrared light. If the samples are optically matched, then tracer particles need to be added. The resolution is much better than the techniques discussed so far, which is particularly interesting to study slip close to the wall. This also counts for the broad range of flow rates that can be probed. The main problem of the technique is the low acquisition rate while for 2D flow, additional detectors need to be implemented.

4 Probing the structure

Generally, shear induced changes in the microstructure of complex fluids underlie changes in the macroscopic rheological response to shear flow. Therefore, to understand the origin of flow instabilities, which are mostly probed on the length scale of tens or hundreds of micrometers,

one has to understand how shear flow affects the structure on the level of the particles that constitute the structure, for which much smaller length scales need to be probed. Even without flow instabilities it is already challenging to understand moderate shear thickening and thinning behavior on the most fundamental microstructural level given the industrial implications of these phenomena.

Ideally a connection can be made between the full 3-D microstructure and the stress trensor, as introduced in chapter c6, on the basis of a microscopic theory which describes this structure. The challenge is to experimentally determine the structure in all three dimensions or at least sufficient measures of the microstructure that can be connected with the measurable components of the stress tensor using theory.

For the conceptually simple case of colloidal spheres the structure is fully determined by the relative position of the particles. In real space this is given by the pair correlation function $g(\mathbf{r})$, which is the probability to find a particle at position $\mathbf{r}$ from the central particle. $g(\mathbf{r})$ is by force obtained by microscopy. Developments in colloid chemistry as well as light microscopy in the eighties and nineties made full characterization of the structure possible. Scattering methods can also be applied to obtain the structure factor $S(\mathbf{q})$ in reciprocal space, which is the Fourier transfer of $g(\mathbf{r})$. Depending on the size of the colloids either visible light or X-rays and neutrons can be used.

For anisotropic particles the situation is more complicated as the structure is not only defined by the relative positions but also by the orientation of the particles. For rods the dependence of the ordering tensor on shear flow can be theoretically derived[50]. The stress tensor is completely determined by the ordering tensor. In the isotropic phase, where there is no orientational ordering nor positional ordering, shear will force the rods to disentangle and align, which results in strong shear thinning. The shear induced (para-)nematic phase might be unstable so that concentration fluctuations set in that can also affect the stability of the flow.

For polymers the situation is even more complicated as position, alignment *and* the shape of the particles themselves are affected by shear flow. Therefore not only the structure needs to be probed but also the shape of the particles. Polymers are generally very small and can only be probed with scattering techniques that use wavelengths in the order of the relevant length scales, see chapters B1 and C1. Generally the q-dependent intensity $I(\mathbf{q})$ is a product of the form factor $P(q)$, which the reciprocal representation of the shape of the particle, and the structure factor $S(\mathbf{q})$, so the both contributions need to be separated. Alternatively, flow birefringent and dichroism measurements can be used to measure changes in the structure and shape of a system as it probes differences in, respectively, the real and imaginary refractive index.

We will focus in this section on the implementation of experiments that probe the structure and shape of the particles *in situ* while the complex fluid is subjected to shear flow. As with the flow profiling, there is a broad range of techniques that can be employed, which partly can be used to do flow profiling. As with flow profiling, it is important in which direction structures are probed and we will see that there are technical limitations to deal with.

4.1 Optical rheometry: polarization

Optical methods can provide important information on the structure of sheared complex fluids, either by using polarimetry or scattering. The *polarization* of light is affected by the material's

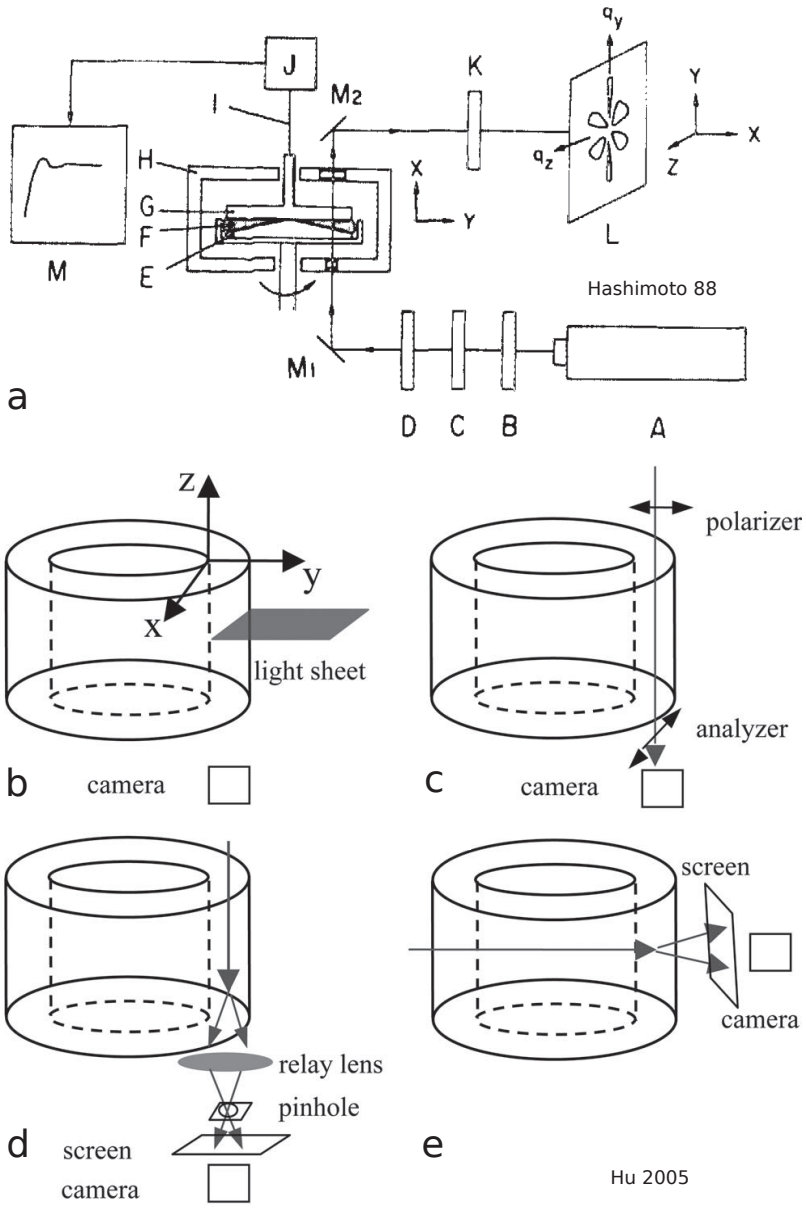


Fig. 6: *Geometries to probes flowing structures with light. (a) A cone-plate shear cell used in SALS experiments probing the flow-vorticity plane. Image taken from Ref. [51]. (b-e) Couette geometry employed in four different experiments, image taken from Ref. [43]. (b) PIV set-up to measure the flow profile; (c) Polarization imaging; (d) scattering directing the beam along the vorticity direction probing the flow-gradient plane in a single plane; (e) scattering directing the beam along the vorticity direction probing the flow-vorticity plane.*

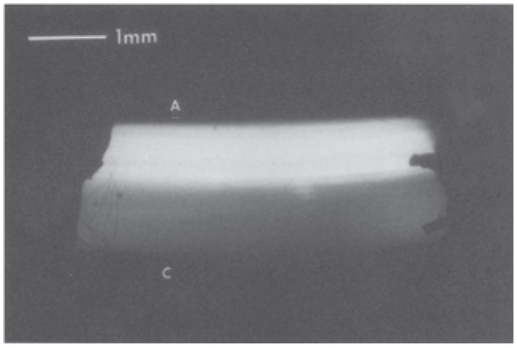


Fig. 7: *Polarization image of a part of a Couette cell loaded with surfactant WLMs at a shear rate of $\dot{\gamma} = 16\ s^{-1}$, using a geometry similar to the one depicted in Ref. 6c. The moving wall is on the inside where the bright birefringent structure can be observed, indicating alignment of the WLMs. Image taken from Ref. [52].*

refractive-index tensor $\mathbf{n} = \mathbf{n}' - i\mathbf{n}''$. The real part $\mathbf{n}'$ induces shifts in the phase of the light, and the imaginary part $\mathbf{n}''$ causes attenuation of its amplitude. Both components can be anisotropic due to shear deformation, with the birefringence $\Delta\mathbf{n}'$ and the dichroism $\Delta\mathbf{n}''$, defined as the differences in the principal eigenvalues of $\mathbf{n}'$ and $\mathbf{n}''$, respectively. The principle eigenvalues are connected to eigenvectors that indicate the angle of the oriented structures. A full account of the use of optics in rheology is given by Fuller [53].
For many polymeric liquids under simple shear, the optical anisotropy that causes birefringence has a simple, linear relationship with components of the stress tensor, known as the stress-optical rule:

$$\mathbf{n}' - n'_m\mathbf{I} = C(\tau - p\mathbf{I}) \ . \tag{1}$$

Here n'_m is the mean refractive index, and p is the hydrostatic pressure. The coefficient C is the stress-optical coefficient. For many polymeric liquids under simple shear, this optical anisotropy has a simple, linear relationship with components of the stress tensor. This reflects the proportionality between stretching of the polymer chains and the stress. The birefringence will, however, saturate at high shear rates, while form effects do not have a linear dependence. Thus, it is difficult to obtain absolute measures of the ordering. In the case of rods, the absolute nematic order parameter S can be simply obtained when the saturation birefringence of perfectly aligned rods, Δn_{sat}, is known:

$$\Delta n = \Delta n_{sat} S \ . \tag{2}$$

Hence, flow birefringence on semi-flexible polymers and WLMs, which can be considered as living semi-flexible polymers, is due to alignment effects of the Kuhn segments, which can be interpreted in terms of sheared rods. There will, however, also be a contribution due to the deformation of the full object.
Nonetheless, flow birefringence is a traditional tool to study flow induced structural changes, as it can be relatively easily implemented in rheological experiments and also be used to characterize flow in complex geometries. As an example, the first experimental proof for the occurrence of structure formation in complex fluids were birefringence measurements showing the split up of a dispersion of WLMs into regions with low and high birefringence, i.e. probably into regions with high and low viscosity, respectively, see Fig. 7 [52]. This strongly suggested that the

shear rate in the two regions is different, which was confirmed by the flow profile experiments we discussed in section 3. Clearly a Couette geometry was used, directing the light through the vorticity direction, so that the flow-gradient plane was probed as shown in Fig. 6c.
With the development of fast acousto-optic modulators, fast simultaneous measurements of birefringence and dichroism are nowadays possible. This was used, for example, to determine the shear-induced shift in isotropic-nematic transition for colloidal rods [54], but also for critical systems of small attractive spheres [55, 56]. Although spheres cannot have any orientation, still the anisotropic flow-induced distortion of the critical structure will render a significant dichroic signal.

4.2 Optical rheometry: light scattering

Scattering of light can also yield important information on sheared complex fluids, especially since the introduction of CCD cameras in the eighties which allowed direct observation of the flow-induced anisotropy in the scattering plane. Structures of about $\Lambda = \frac{2\pi}{q} = \frac{\lambda}{2\sin\theta_s/2} \approx 350\ nm$ or bigger can be observed, assuming that the scattering angle θ_s should be not more than about 90 degrees for planar detection and using a wavelength of 500 nm. This means that this Small Angle Light Scattering (SALS) is suited for relatively big spherical colloids but also to observe density fluctuations in complex fluids consisting of smaller particles. A typical example of scattering on a single particle level dates back to Ackerson and co-workers who studied the shear-induced distortion of colloidal crystals [59, 60, 61]. A typical example of scattering from density fluctuations is due to Dhont and co-workers, who studied the shear induced deformation of the critical structure factor of small attractive spheres close to the gas-liquid phase transition [62, 63].
Light scattering has also been applied for very slender particles like WLMs and polymer blends. Hashimoto et al were the first to observe very anisotropic "Butterfly" scatter patterns for sheared polymer solutions [51, 57], as displayed in 8a. For this study a cone-plate geometry was used where the light was directed along the gradient direction, probing the flow-vorticity scattering plane, see Fig. 6b. As the constituting polymers are much too small to scatter light in the observable q-range, these patterns suggest that the scattering is due to concentration fluctuations.
Similar patterns have been observed for shear banding WLMs [43]. In this paper all optical tools that we discussed so far were used: PIV to obtain velocity profiles, Fig. 6b, polarization to obtain the location of more or less oriented structures, Fig. 6c, and light scattering in two scattering geometries: in the radial direction of a Couette cell, where light is directed through the gradient direction and the flow-vorticity plane is probed, Fig. 6d, and directing the light along the vorticity direction, to take scatter patterns in the flow-gradient plane, Fig. 6e. The latter scattering geometry has the problem that scattering comes from the full length of the Couette cell, which smears out the angular resolution. Hu and Lips employed, however, a pinhole in the detection line to ensure that the observed scattering comes from a plane. The other problem is that the long pathway of the light through the sample will lead to multiple scattering so that only slightly turbid samples can be used. The advantage of this scattering geometry is twofold. For many systems the eigen vector belonging to the principle eigen value that describes the ordering in the system is located in the flow-gradient plane, so that the full information is obtained when probing this plane. Moreover, the gap of the shear cell can be scanned, which is important when gradient bands form. Instead of adding up the contribution

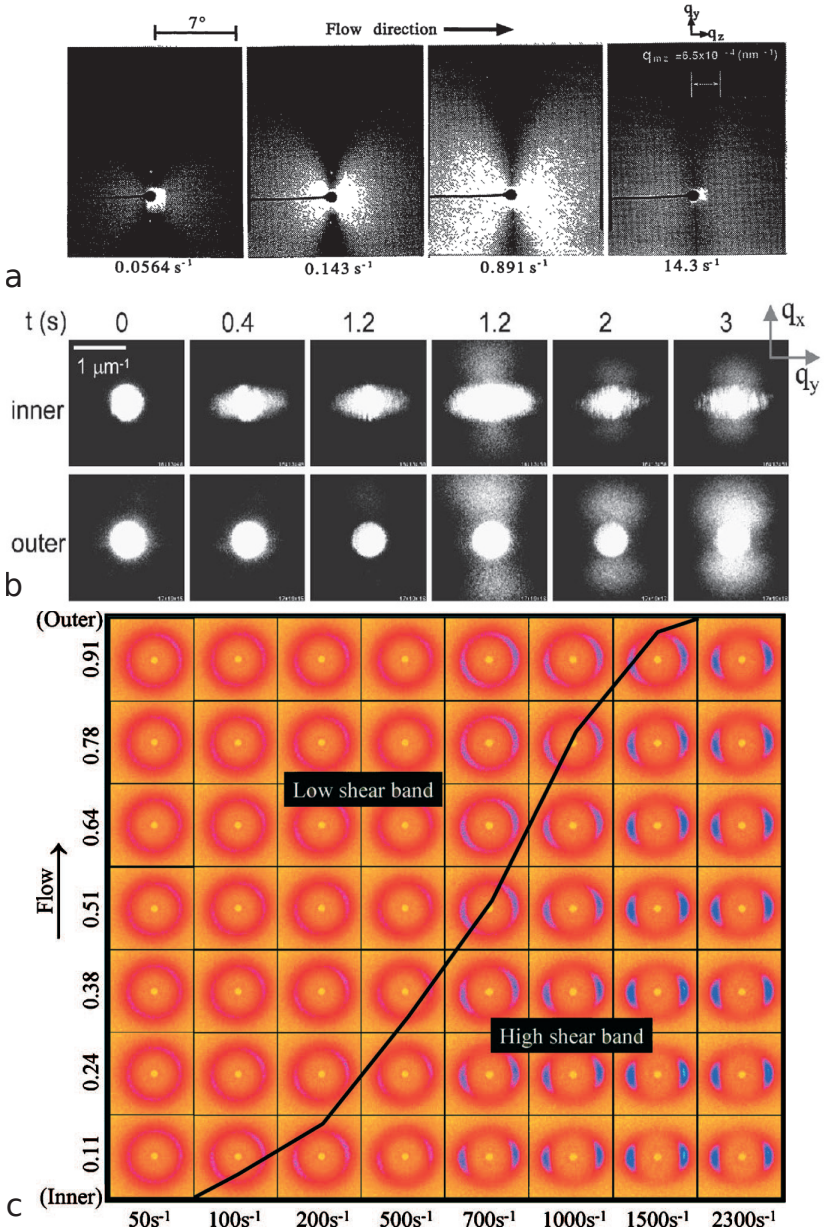


Fig. 8: *SALS (a,b) and SANS (c) patterns of a polymer solution (a, image taken from Ref. [57]) and surfactant WLMs (b,c). The shear patterns of a sheared polysteren solution in (a) are taken using the set-up displayed in Fig. 6a. (b) Time evolution of SALS patterns from the inner and outer side of the gap in the flow-gradient plane after applying a shear rate of* $13\ s^{-1}$ *to a WLM solution, using the geometry depicted in 6d. Image taken from Ref. [43]. (c) SANS patterns of a sheared CTAB solution taken at different shear rates and different positions throughout the gap of the shear cell, see Fig. 10a. Note that the SALS data in b all fall within the q-range of the beam stop in the center of the SANS data in c. Image taken from Ref. [58].*

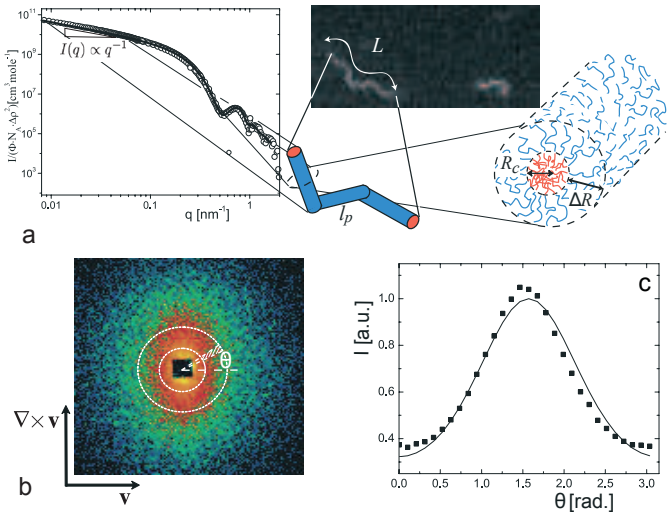


Fig. 9: *(a) Block-copolymer WLM at all relevant length scales: Angle-averaged SANS curve at zero shear on the left.* $I(q) \propto q^{-1}$ *at low* q*. Here the stiff Kuhn segments, characterized by the persistence length* l_p*, are probed. The cross-section of the micelle,characterized by the shell thickness* ΔR *and the core radius* R_c *are obtained from the structure at high* q*. The fluorescence microscopy image of a block-copolymer WLM on the right gives the worms contour length* L*. (b) Scattering pattern of the same WLM at a volume fraction of 1% at a shear rate of* $\dot{\gamma} = 1s^1$ *in the flow-velocity plane. The dashed lines indicate the* q*-range that is used to obtain the azimuthal intensity profile as plotted in (c). Here* θ *is the angle with the shear flow and the full line indicates a fit to Eq. 3. Images taken from Refs. [64, 65].*

of bands with different structure, which is the case when directing light through the gradient direction, see Fig. 2b, the isolated contributions of the bands can be probed. This can be seen in Fig. 8b, where the development of the SALS patterns after applying a shear rate into the shear banding region is displayed close to the inner and outer wall.
The origin of these structures is not yet clear, but certainly the size is an order of magnitude bigger than the Kuhn segments of which the ordering was probed by the polarization experiments. In order to probe this length scale either X-rays or neutrons need to be used, which are of the order of $1\ nm$, which we will treat in the next two sections.

4.3 *In situ* Scattering: Neutrons

As explained in chapters B1 and C1, Small Angle Neutron Scattering is an excellent technique to probe soft materials such as polymers and proteins. Although the neutron sources are extremely weak in terms of flux as compared to X-ray sources, neutrons have the great advantage that, in combination with chemistry, scattering contrast can be reduced to zero so that particles can be made invisible. Thus, a clear distinction can be made between the effect of shear flow on the form of particles, given by the form factor $P(q)$ and the structure of the fluid, given by the structure factor $S(\mathbf{q})$, by either making all particles visible for neutrons or only a fraction. An

additional advantage is that the scattering cross section can be exactly calculated, knowing the composition of the fluids, which guarantees also a perfect reduction of the background scattering. This is, for example, very helpful when the orientational distribution of Kuhn segments is obtained from the scatter pattern.

In Fig. 9 it is shown how from the scattered 2D profile the orientational distribution can be obtained. In this example WLMs consisting of block-copolymers were used, which can also be fluorescently labeled and imaged with fluorescent microscopy, as shown on the right of Fig. 9a [64]. The core and shell radii and the aggregation number per unit length can be obtained from the high q-range of the intensity profile, as shown in Fig. 9a. This nano-structure of the WLM is not affected by shear flow. At the low q-range, the scattered intensity is proportional to $I \sim q^{-1}$, which is typical for scattering of rods. In Fig. 9b this q-range is indicated by the two circles for a typical scatter pattern of block-copolymers WLMs under shear conditions ($\dot{\gamma} = 1\ s^{-1}$), which clearly hints to an anisotropic distribution of Kuhn-segments. The orientational distribution of the rods is proportional to the radial averaged azimuthal profile, see Fig. 9c. Assuming a Maier-Saupe type of orientation distribution function, the azimuthal intensity profile from the ordered phase $I(q, \theta)$ is generally well described by

$$I(q, \theta) = I_0 \exp[\alpha P_2(\theta) - 1], \tag{3}$$

where the parameter α describes the width of the intensity profile. $\langle P_2(\theta) \rangle$, which is the same as the order parameter S in Eq. 2 for the polarization experiment, can then be easily calculated using

$$\langle P_2(\theta) \rangle = \frac{\int_0^\pi \exp[\alpha P_2(\theta)] P_2(\theta) \sin(\theta) d\theta}{\int_0^\pi \exp[\alpha P_2(\theta)] \sin(\theta) d\theta}. \tag{4}$$

Thus, in SANS experiments the ordering can be unambiguously attributed to the Kuhn segments, without any calibration or assumption.

The orientation of Kuhn segments of surfactant WLMs was first studied using SANS by Thurn et al. [66]. Later, SANS was used to estimate the shear-induced isotropic-nematic phase transition [67]. There are, however, two experimental problems with these measurements. First, the data is taken from a shear banding system, directing the beam through the gradient direction. This means that contributions from the high and low aligned bands are added, as explained in section 4.2. Second, the main disadvantage of SANS is the very low flux of the source, despite of recent developments in using spallation sources. As a consequence it is most suited to probe steady states, but, as we learned in section 3, steady states can be rare in complex flows, while the microscopic origin of flow instabilities are best studied using transient experiments. Both of these problems were recently addressed. We will first show how the issue of the low flux can be solved and then discuss developments in the shear geometries that solve the issue of the averaging over the gap.

In order to observe transient behavior in sheared complex fluids with neutrons, a repetitive flow field needs to be employed so that shear deformation can be linked to a stroboscopic collection of neutrons. When an electronic trigger supplied by the rheometer initiates the collection of data from the scattered neutrons, then $n \leq 400$ SANS patterns of time resolution $\Delta t = (n\omega/2\pi) - 1$, where $\omega/2\pi$ is the frequency of the applied excitation field, are summed in time bins with respect to the trigger. To get enough counting statistics for each time channel the histograms of many shear cycles were summed up over a quarter of an hour to hours.

This technique has been employed in combination with Large Amplitude Oscillatory Shear flow (LAOS), where a time-dependent sinusoidal shear rate is applied to the fluid [64, 68]. The response of block-copolymer WLMs, which are known to be stable entities, and its dependence on the concentration could be related to the mechanical response, using Smoluchowski theory for rods [64], despite of the polydispersity and relative flexibility. In this way the isotropic-nematic spinodal point could be identified. For the same system flow profiling with heterodyne light scattering has been performed, see Fig. 3a, where shear bands indeed have been found close to the spinodal point as predicted by theory [6]. This work thus shows the relation between the microscopic and macroscopic response and resulting flow instability, as we set out to do. The response of surfactant WLMs can be markedly different, probably due to the living character of these entities. For this system it could be shown that during a cycle a transformation takes place from a solid-like response to a fluid-like response after it yields into a shear banding state. At this yielding point, the Kuhn segments display a much higher orientation, they over-orient, compared to stationary flow. Moreover, a stressorientation rule which is rate-based at low frequencies and strain-based at high frequencies could be identified.
As we mentioned before, all of these experiments suffer from the fact that averages are taken over the gap, where different bands might co-exist. This problem has been solved by the group of Wagner, where a Couette shear cell was developed, that is oriented with the vorticity axis in the direction of the beam, as shown in Fig. 10a, so that again the flow-gradient plane can be probed and even scanned, using small slits [69]. In this way the degree of ordering as well as the orientation of the principle eigen vector was found and not merely the projection in the flow-vorticity plane as in the experiments described before. Moreover, the structure could be probed throughout the gap, as can be seen in Fig. 8c. Thus, a qualitative correspondence between the orientational ordering of Kuhn segments and macroscopic rheological stress has been found under steady-shear conditions by linking rheo-SANS measurements to the Giesekus model [58, 70].
This set-up has the disadvantage that no on-line rheology can be done and the flux, which is anyway small for neutrons, is heavily reduced. Moreover, for SANS experiments it is impossible to obtain information of the local velocity, so that additional flow profiling needs to be done. In the next section we will show how X-rays can, in principle, solve all of these issues.

4.4 *In situ* Scattering: X-rays

Small Angle X-ray Scattering (SAXS) is especially useful to obtain structures of particles that contain atoms with high electron densities. As such SAXS is less suited to study WLMs. The main advantage of SAXS is the incredibly high flux due to the synchrotron sources, which does bare, however, the problem of beam damage. The high flux allows for very local probing with beam diameters up to 1 μm radius and fast probing. The latter also depends on the detectors, but nowadays X-ray sensitive CCD cameras are available that can make up to 5000 frames per second with a single photon precisions, which means that almost every scattered photon that hits the chip will be detected.
SAXS has mostly been combined with Couette cells, as for SANS, but as with SANS, shear cells have been developed to obtain the scattering in the flow gradient plane. Caputo and Burghardt

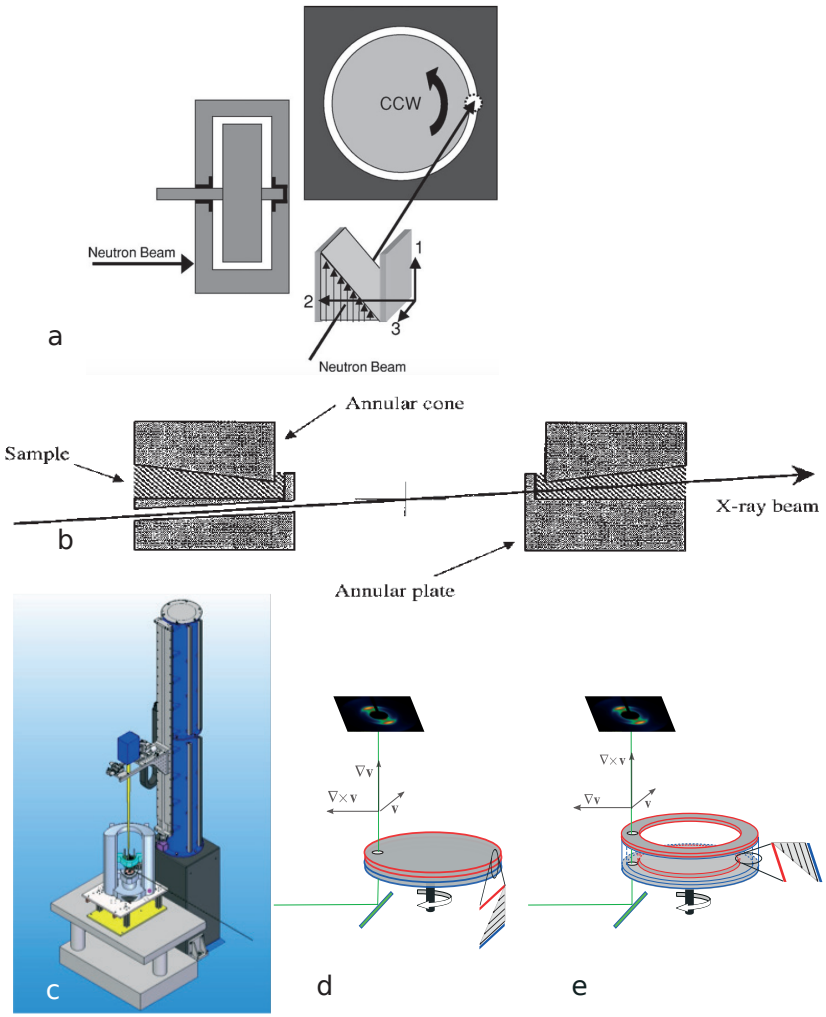


Fig. 10: *(a) SANS set-up directing the beam along the vorticity of a tilted Couette cell which is hermetically closed. Image taken from Ref. [69]. (b) SAXS set-up directing the beam along the vorticity of a tilted cone-plate set-up. Image taken from Ref. [71]. (c) SAXS set-up where the beam is reflected into the vertical direction using a crystal. Image taken from Ref. [72]. This set-up can be either used with a plate/cone-plate geometry (d) or with a Couette geometry (e), probing either the flow-vorticity or flow-gradient plane, respectively.*

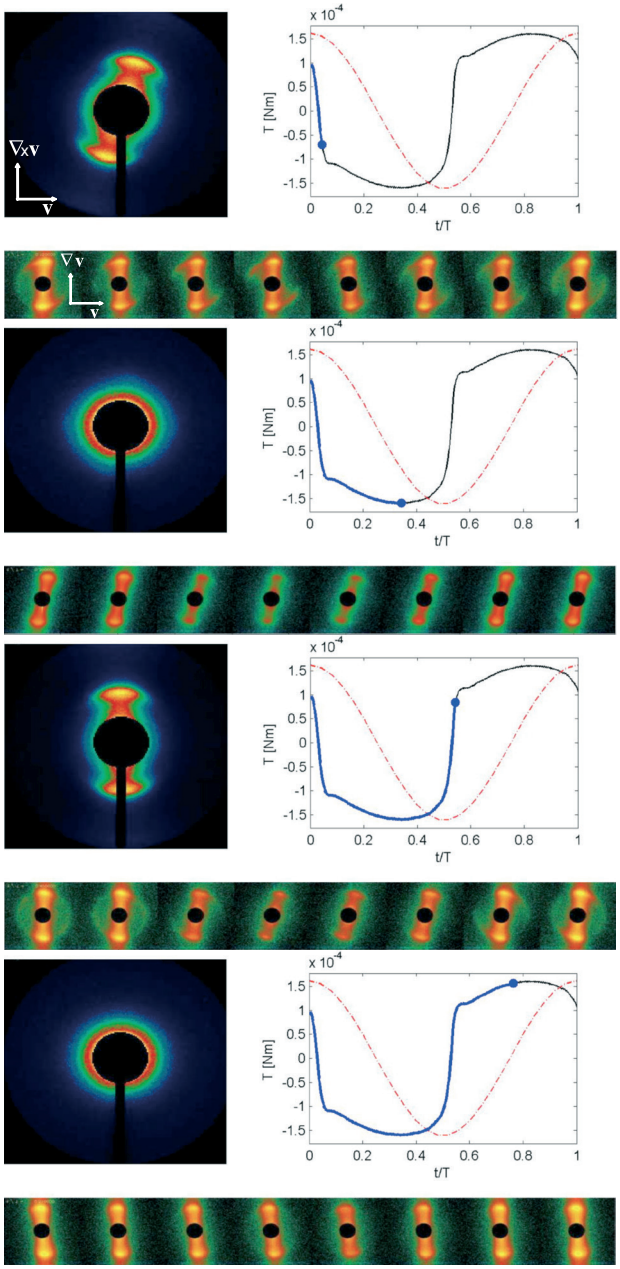


Fig. 11: *Four snap shots of a nematic dispersion of colloidal platelets subjected to LAOS at a frequency of* $0.04Hz$ *and a strain amplitude of* $\gamma_{max} = 12.8$*. The single pattern in the flow-vorticity plane is taken using the plate-plate geometry depicted in Fig. 10d; the row of patterns in the flow-gradient plane is taken using the Couette geometry depicted in Fig. 10e. The patterns are taken in the 1 mm gap between the moving wall at the left and the standing wall on the right. The two curves show the stress response (blue) and applied strain (red). The dot indicates the point during the oscillation where the patterns are taken.*

build a cone-plate cell which was slightly tilted so that the beam could be directed along the vorticity direction, as can be seen in Fig. 6 [71]. They used this cell to observe the tumbling behavior of dense lyotropic liquid crystalline polymers, which can only be observed in the flow-gradient plane.
Similar tumbling behavior has been observed for the nematic phase of colloidal platelets subjected to Large Oscillatory Shear Flow [73]. For this study a modified rheometer was combined with a Small-Angle X-ray set-up where the X-rays are deflected in the vertical direction, as shown in Fig. 10c, thus passing through the gradient direction of a plate-plate geometry as shown in Fig. 10d [72, 74, 73]. The use of a plate-plate geometry is especially advantageous for studies on polymer melts, since these systems can hardly be loaded in the Couette cells that are generally used for *in situ* scattering experiment. This set-up is however especially useful in combination with the Couette geometry, which can easily be implemented. The beam can be directed along the vorticity direction, while, in contrast with the other set-ups we discussed above, it is still possible to do rheology. Moreover, with the small diameter of the beam it is possible to scan the gap with a high spacial resolution. Of course the flux is significantly reduced both due to the reflection of the beam and by squeezing the beam, but with the present fluxes at the modern synchrotrons guarantee that SAXS experiments can be done at a high time resolution.
Using the combination of both geometries we could show how this system undergoes a transition during a large amplitude oscillation from an elastic to a fluid-like response. The intensity peaks in the scatter patterns shown in Fig 11 are due to the inter-particle scattering between the faces of the oriented platelets. As rheology and SAXS are synchronized, we observe that the fluid yields after a stress overshoot at the point where the platelets flip their orientation in both observable directions. The SAXS scan of the gap also shows, however, that this response is very inhomogeneous throughout the gap. The ordering is higher close to the wall. This effect is of course overseen in the plate-plate geometry because here the scattering is averaged over the gap. Hence, we learn that it is very important to probe the structure throughout the gap, which is possible due to the tiny beam diameter.
With this experiment we simultaneously obtain the structure, spatially resolved, and the rheology. It is in principle possible to obtain even the flow profile using X-rays. The nowadays available high flux and brilliance of third generation synchrotron sources can be exploited to generate coherent X-rays and access the dynamical behavior, as introduced in section 3.1. In an X-ray Photon Correlation Spectroscopy (XPCS) experiment, motion of scatterers leads to changes of the interference pattern, which are quantified by intensity auto-correlation functions [75, 76]. Recently XPCS has been applied to measure local velocities in microfluidics [77]. The only problem is that the fluid needs to contain strong scatterers. If this is the case then, in principle, it is possible to obtain microstructure, the macroscopic stress response and the flow profile in one single experiment.

4.5 *In situ* Microscopy

Despite of the rich information that is obtained by the manifold of experiments described above on the semi-flexible WLMs, still there are many open questions. For example, the origin of the shear thinning and the role of the 'living' character of the WLMs and their stiffness is still not clarified. This is partly due to the fact that all information was obtained in reciprocal space. It is

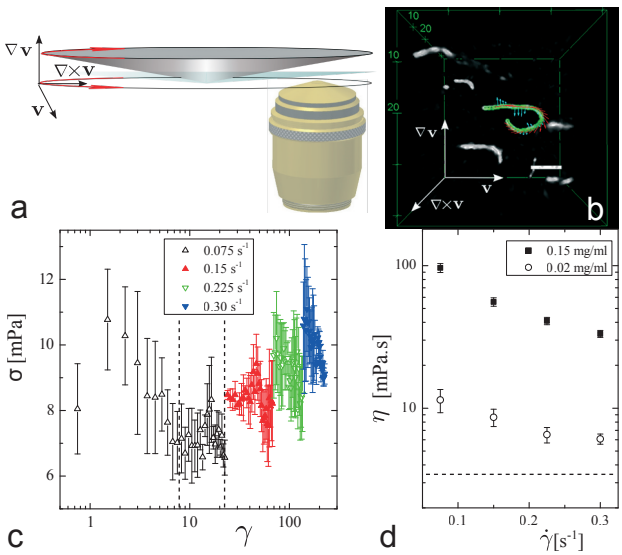


Fig. 12: *(a) Schematic depiction of the counter-rotating cone-plate shear cell placed on an inverted microscope. (b) Typical reconstructed 3D stack of confocal images, showing a small fraction of labeled actin filaments embedded in a dark background of unlabeled F-actin. Green line shows a tracked filament and the small blue and green arrows indicate the local tangent and binormal vector. Scale bar:* $10\,\mu m$*, tick unit:* μm*. (c) Stress as a function of dimensionless strain where the shear rate is increased every five minutes at a concentration of* c_{high}*= 0.15 mg/ml. (d) Steady-state viscosity as obtained from the strain window where the stress is constant (see dashed lines in (c)) as a function of shear rate for the two concentrations. The dashed line gives the solvent viscosity. Image taken from Ref. [78].*

therefore a challenge to find systems that can be imaged in real space using the newly available fast confocal microscopes. There are two requirements. First, the systems needs to be suited. It needs to be labeled and stiff enough so that they do not appear as a coil. This means that the persistence length needs to be at least of the order of a micrometer, given the resolution of a confocal microscope of about $350\ nm$. In Fig. 9a we have shown an image of a fluorescently labeled block-copolymer WLM, for which the persistence length is however too small to obtain a full 3-D contour. Second, a set-up is needed where the structures do not move out of the field of view during the acquisition. Such a set-up is depicted in Fig. 12a, showing a cone-plate geometry where cone and plate can counter rotate. As a consequence there is a zero velocity plane where structures are sheared but do not move out of the field of view.
This set-up was used to study dispersions of F-actin, which are stiff filaments with a persistence length of about $17\ \mu m$ [78]. In Fig. 12b a reconstructed 3D image of labeled filaments in a background of unlabeled filaments is shown. The problem with imaging is always how to transform images in useful numbers that describe the system. In this case it was obvious that highly entangled filaments completely disentangle when subjected to shear flow, forming loops in the flow-vorticity plane that slide over each other. This microscopic behavior causes the strain softening depicted in Fig. 12c, where it can be seen that the stress diminishes between strain $\gamma = 1$ and $\gamma = 10$. The statistical proof was given by the sharp distribution that was obtained of the local binormal vectors for the parts of the segments where the filament is highly bended, showing that the loops in the filaments are always oriented with the normal in the gradient direction. At high shear rates it can be observed that the stretched parts of the filaments become more oriented, shown by the distribution of the tangents, which explains the shear thinning shown in Fig. 12d. Thus, these real space experiments give a full microscopic insight in the origin of strain softening and shear thinning for an important class of systems. What was not reported, but what could have been done, was the flow rate throughout the gap of the cell. In principle a flow profile can be taken if the gap of the shear cell at the location of the objective is not bigger than the working distance of the objective. This has indeed been done for dispersions of colloidal spheres [79, 80, 81].

5 Conclusions

It is clear that a huge effort is being undertaken to fulfill the requirements for understanding the flow of complex fluids, which is to measure mechanical responses, as well as structural responses and flow profiles for one and the same system at the same time. We have introduced many techniques which fulfill part of the requirements, or even all requirements but only for a limited class of samples. Most of the techniques are still in full development, so that this overview will soon be outdated. Thus, working with complex fluids and complex flow it is important to stay informed. Moreover, it is important to pick the right technique for the problem of interest. We hope that this chapter has raised the curiosity of the reader and will aid in the quest of interesting complex fluids and suited techniques to study them.

References

[1] S. Manneville, Rheologica Acta **47**(3), 301 (2008).

[2] G. C. Maitland, Curr. Opinion Colloid Interface Sci. **5**, 301 (2000).

[3] M. J. Decker, C. J. Halbach, C. H. Nam, N. J. Wagner, and E. D. Wetzel, Composites Sci. and Tech. **67**, 565 (2007).

[4] P. D. Olmsted and C. Y. D. Lu, Phys. Rev. E **60**(4), 4397 (1999).

[5] P. D. Olmsted, Europhys. Lett. **48**, 399 (1999).

[6] J. K. G. Dhont, Phys. Rev. E **60**(4), 4534 (1999).

[7] J. K. G. Dhont and W. J. Briels, Rheo. Acta **47**(3), 257 (2008).

[8] P. D. Olmsted, Rheol. Acta **47**(3), 283 (2008).

[9] A. Imhof, A. van Blaaderen, and J. K. G. Dhont, Langmuir **10**, 3477 (1994).

[10] W. M. Holmes, P. T. Callaghan, D. Vlassopoulos, and J. Roovers, J. Rheol. **48**(5), 1085 (2004).

[11] J. Sprakel, E. Spruijt, M. A. C. Stuart, N. A. M. Besseling, M. P. Lettinga, and J. van der Gucht, Soft Matter **4**(8), 1696 (2008).

[12] P. Tapadia and S. Q. Wang, Physical Review Letters **91**(19) (2003).

[13] P. Tapadia, S. Ravindranath, and S. Q. Wang, Physical Review Letters **96**(19) (2006).

[14] P. E. Boukany, Y. T. Hu, and S. Q. Wang, Macromolecules **41**(7), 2644 (2008).

[15] Y. T. Hu, C. Palla, and A. Lips, Macromolecules **41**, 6618 (2008).

[16] N. A. Spenley, M. E. Cates, and T. C. B. McLeish, Phys. Rev. Lett. **71**, 939 (1993).

[17] M. E. Cates, Macromolecules **20**(9), 2289 (1987).

[18] H. Rehage and H. Hoffmann, J. Phys. Chem. **92**, 4712 (1988).

[19] H. Rehage and H. Hoffmann, Mol. Phys. **74**, 933 (1991).

[20] M. A. Fardin, T. J. Ober, V. Grenard, T. Divoux, S. Manneville, G. H. McKinley, and S. Lerouge, Soft Matter **8**(39), 10072 (2012).

[21] P. Boltenhagen, Y. Hu, E. F. Matthys, and D. J. Pine, Phys. Rev. Lett. **79**(12), 2359 (1997).

[22] R. G. Larson, E. S. G. Shaqfeh, and S. J. Muller, J. Fluid Mech. **218**, 573 (1990).

[23] S. Lin-Gibson, J. A. Pathak, E. A. Grulke, H. Wang, and E. K. Hobbie, Physical Review Letters **92**(4) (2004).

[24] F. Pignon, A. Magnin, and J. M. Piau, Phys. Rev. Lett. **79**(23), 4689 (1997).

[25] R. H. Tromp and E. H. A. D. Hoog, Phys. Rev. E **77**, 031503 (2008).

[26] K. Kang, M. P. Lettinga, Z. Dogic, and J. K. G. Dhont, Phys. Rev. E **74**, 026307 (2006).

[27] P. Pakdel and G. H. McKinley, Phys. Rev. Lett. **77**(12), 2459 (1996).

[28] M. A. Fardin, D. Lopez, J. Croso, G. Grégoire, O. Cardoso, G. H. McKinley, and S. Lerouge, Phys. Rev. Lett. **104**, 178303 (2010).

[29] M. A. Fardin, T. J. Ober, C. Gay, G. Gregoire, G. H. McKinley, and S. Lerouge, Soft Matter **8**(4), 910 (2012).

[30] S. M. Fielding, Phys.Rev. E **76**, 016311 (2007).

[31] R. G. Larson, Rheol. Acta **31**, 213 (1992).

[32] J. Delgado, H. Kriegs, and R. Castillo, Journal of Physical Chemistry B **113**(47), 15485 (2009).

[33] T. Gallot, C. Perge, V. Grenard, M. A. Fardin, N. Taberlet, and S. Manneville, Review of Scientific Instruments **84**(4) (2013).

[34] D. P. Chowdhury, C. M. Sorensen, T. W. Taylor, J. F. Merklin, and T. W. Lester, Applied Optics **23**(22), 4149 (1984).

[35] S. Busch, T. H. Jensen, Y. Chushkin, and A. Fluerasu, European Physical Journal E **26**(1-2), 55 (2008).

[36] J. B. Salmon, A. Colin, S. Manneville, and F. Molino, Physical Review Letters **90**(22) (2003).

[37] P. D. Olmsted and C. Y. D. Lu, Phys. Rev. A **56**, 55 (1997).

[38] P. T. Callaghan, M. E. Cates, C. J. Rofe, and J. B. A. F. Smeulders, J. Phys. II France **6**, 375 (1996).

[39] E. Miller and J. P. Rothstein, J. Non-Newt. Fluid **143**(1), 22 (2007).

[40] M. P. Lettinga and S. Manneville, Physical Review Letters **103**(24) (2009).

[41] M. M. Britton and P. T. Callaghan, Physical Review Letters **78**(26), 4930 (1997).

[42] M. A. Fardin and S. Lerouge, European Physical Journal E **35**(9) (2012).

[43] Y. T. Hu and A. Lips, J. Rheol. **49**(5), 1001 (2005).

[44] S. Manneville, L. Becu, and A. Colin, European Physical Journal-Applied Physics **28**(3), 361 (2004).

[45] R. Angelico, S. Amin, M. Monduzzi, S. Murgia, U. Olsson, and G. Palazzo, Soft Matter **8**(42), 10941 (2012).

[46] P. T. Callaghan, Rheologica Acta **47**(3), 243 (2008).

[47] M. Harvey and T. A. Waigh, Physical Review E **83**(3) (2011).

[48] S. Jaradat, M. Harvey, and T. A. Waigh, Soft Matter **8**(46), 11677 (2012).

[49] A. V. Malm, A. W. Harrison, and T. A. Waigh, Soft Matter **10**(41), 8210 (2014).

[50] J. K. G. Dhont and W. J. Briels, Colloid Surface A **213**(2-3), 131 (2003).

[51] T. Hashimoto, T. Takebe, and S. Suehiro, Journal of Chemical Physics **88**(9), 5874 (1988).

[52] J. P. Decruppe, R. Cressely, R. Makhloufi, and E. Cappelaere, Colloid Polym. Sci. **273**(4), 346 (1995).

[53] G. G. Fuller, Annual Review of Fluid Mechanics **22**, 387 (1990).

[54] T. A. J. Lenstra, Z. Dogic, and J. K. G. Dhont, Journal of Chemical Physics **114**(22), 10151 (2001).

[55] T. A. J. Lenstra and J. K. G. Dhont, Physical Review E **63**(6), art. no. (2001).

[56] J. K. G. Dhont and H. Verduin, J. Chem. Phys. **101**(7), 6193 (1994).

[57] T. Hashimoto and T. Kume, Journal of the Physical Society of Japan **61**(6), 1839 (1992).

[58] M. E. Helgeson, P. A. Vasquez, E. W. Kaler, and N. J. Wagner, J. Rheol. **53**(3), 727 (2009).

[59] B. J. Ackerson and N. A. Clark, Physical Review Letters **46**(2), 123 (1981).

[60] B. J. Ackerson and P. N. Pusey, Phys. Rev. Lett. **61**, 1033 (1988).

[61] B. J. Ackerson, Journal of Rheology **34**(4), 553 (1990).

[62] M. P. Lettinga, H. Wang, and J. K. G. Dhont, Phys. Rev. E **70**, 061405 (2004).

[63] J. K. G. Dhont and I. Bodnar, Physical Review E **58**(4), 4783 (1998).

[64] B. Lonetti, J. Kohlbrecher, L. Willner, J. K. G. Dhont, and M. P. Lettinga, J. Phys.: Condens. Matter **20**, 404207 (2008).

[65] B. Lonetti, A. Tsigkri, P. R. Lang, J. Stellbrink, L. Willner, J. Kohlbrecher, and M. P. Lettinga, Macromolecules **44**(9), 3583 (2011).

[66] H. Thurn, J. Kalus, and H. Hoffmann, Journal of Chemical Physics **80**(7), 3440 (1984).

[67] J. F. Berret, D. C. Roux, and P. Lindner, Eur. Phys. J. B **5**, 67 (1998).

[68] S. A. Rogers, J. Rheol. **56**(5), 1129 (2012).

[69] M. W. Liberatore, F. Nettesheim, N. J. Wagner, and L. Porcar, Physical Review E **73**(2) (2006).

[70] M. E. Helgeson, M. D. Reichert, Y. T. Hu, and N. J. Wagner, Soft Matter **5**, 3858 (2009).

[71] F. E. Caputo and W. R. Burghardt, Macromolecules **34**(19), 6684 (2001).

[72] B. Struth, K. Hyun, E. Kats, T. Meins, M. Walther, M. Wilhelm, and G. G., Langmuir **27**, 2880 (2011).

[73] M. P. Lettinga, P. Holmqvist, P. Ballesta, S. Rogers, D. Kleshchanok, and B. Struth, Phys. Rev. Lett. **109**, 246001 (2012).

[74] D. Denisov, M. T. Dang, B. Struth, G. Wegdam, and P. Schall, Scientific Reports **3** (2013).

[75] M. Sutton, S. G. J. Mochrie, T. Greytak, S. E. Nagler, L. E. Berman, G. A. Held, and G. B. Stephenson, Nature **352**(6336), 608 (1991).

[76] T. ThurnAlbrecht, W. Steffen, A. Patkowski, G. Meier, E. W. Fischer, G. Grubel, and D. L. Abernathy, Physical Review Letters **77**(27), 5437 (1996).

[77] A. Fluerasu, P. Kwasniewski, C. Caronna, F. Destremaut, J. B. Salmon, and A. Madsen, New Journal of Physics **12** (2010).

[78] I. Kirchenbuechler, D. Guu, N. A. Kurniawan, G. H. Koenderink, and M. P. Lettinga, Nature Communications **5** (2014).

[79] D. Derks, H. Wisman, A. van Blaaderen, and A. Imhof, J. Phys.: Condens. Matter **16**, S3917 (2004).

[80] I. Cohen, B. Davidovitch, A. B. Schofield, M. P. Brenner, and D. A. Weitz, Physical Review Letters **97**(21) (2006).

[81] R. Besseling, L. Isa, P. Ballesta, G. Petekidis, M. E. Cates, and W. C. K. Poon, Physical Review Letters **105**(26) (2010).

F 2 High Pressure

G. Meier
Institute of Complex Systems, ICS-3
Forschungszentrum Jülich GmbH

Contents

1 Introduction 2

2 Static and Dynamic Pressure Effects in Soft Matter 3

2.1. Pressure Effects on Structure in Biological Systems 3
2.2. Pressure Jump Techniques to Study Phase Behavior in Colloidal Sytems 7

3 High Pressure Microscopy 10

3.1 Pressure Cells for Microscopes 10
3.2 Soft Matter Science Applications 12

References 14

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

Pressure is a frequently neglected state variable, which gains most of its importance today in possible technical applications. However, on earth mainly two different scenarios are notable, first the pressures which are achievable in deep sea. In this environment pressures up to 1.000bar or equivalently 100MPa, (1bar=100kPa=10^5 N/m^2=10^5 kg/ms^2) corresponding to about 10.000m water depth are reached. The other scenario is determined by the enormous pressures occurring in the world's interior. Here pressures up to 364GPa are reached.

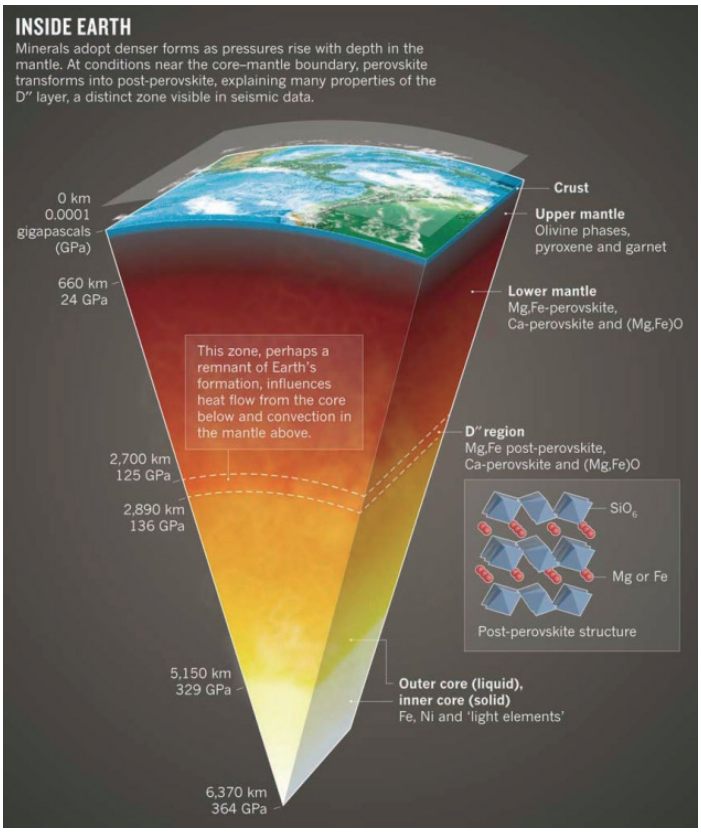


Fig. 1: *Pressure variation inside the earth. Near the centre of the core pressures up to 364 GPa are reached. The increase of pressure with distance from the mantle surface is roughly linear and amounts to about 500 bar/km. Likewise the temperature increases by about 25K/km.*

Clearly these pressures induce phase transitions between the various materials, which form the earth's mantle and core. In Fig.1 this scenario is schematically depicted. In this lecture this physics with its geologically interesting phenomena will not be discussed. We are rather concerned with soft matter physics, which takes place in the above mentioned pressure range of about 100-500MPa (1000-5000bar). In this above mentioned pressure range biological processes can occur and hence this range is the natural one with regard to bio-soft scientific investigations. The influence of high pressure on biological systems was first described more than one hundred years ago as a method to preserve milk. Recent investigations with high

pressure have been undertaken to inactivate microorganisms that spoil food, to sterilize and pasteurize foods as well as pharmaceutical products. Treatment with high pressure is a powerful tool, because microbial inactivation is not associated with as many unwanted chemical reactions as those that occur during thermal treatment. By this degradation would take place much more easily [1]. A scientific interesting example is the formation of early life forms in the vicinity of black smokers in the deep sea at high pressures [2].
Another technologically important field is the solubilisation of crude oil in large sea depths and the storage of carbon dioxide in the deep sea. The continuously increasing world demand for petroleum forces the oil industry to exploit reservoirs at larger depth (off-shore), which means at conditions of high pressure. In oil reservoir engineering, the crucial data needed to determine the economic exploitability of a hydrocarbon reservoir are: pressure, gas-oil contact, rock porosity, oil saturation and fluid properties (such as oil composition, density and viscosity). Currently, this process is done with the aid of reservoir simulators, which require knowledge of the initial state of the natural hydrocarbon reservoirs, and a model for the compositional variation. To have access to experiments, which mimic deep sea conditions on-line in the lab, would greatly help in optimizing exploitation [3].
Tribology (which deals with interacting surfaces in relative motion) is also important to understand from the engineer's point of view and has a large technological impact. Here the knowledge of transport properties of soft matter like the viscosity of grease in gearings is of great importance, albeit this quantity is difficult to obtain experimentally [4].
Last but not least, high pressure induced chemical reactions are largely used in production, may be the most prominent one is the synthesis of ammonia by the famous Haber-Bosch process [5].

2 Static and Dynamic Pressure Effects in Soft Matter

2.1. Pressure Effects on Structure in Biological Systems

We will, in this lecture, concentrate on mainly biological applications as these undergo recently are tremendous development. While for small and simple systems pressure effects are often disappointingly small, they can reach large and demanding effects for complex large biological molecules. This is intrinsically interesting as non-covalent interactions dominate structure and function of these items. The complex structures are easily subjected to pressure effects due to the subtle equilibrium between the different interacting components. Biological functionality is often strongly governed by a structural prerequisite, which in turn can be understood by these equilibrium conditions. If cavity effects are present being part of the structure then pressure has a large field of operation and interesting phenomena are to be expected. Naturally all spectroscopic tools yielding structural information about biologically relevant complex structures have been employed, X-ray scattering, light scattering, neutron scattering, NMR investigations, infrared- to florescence spectroscopy, microscopy in different arrangements like for example confocal imaging [6]. In principle static structural investigations are possible but more interestingly also structural developments initiated by a pressure jump. This is part of the beauty of the method because a temperature jump is much less decisive. Let us start with some examples on structural investigations, where the application of pressure is an interesting issue.
In a SANS study of the glucose/xylose isomerase prepared from Streptomyces Rubiginosus we were interested in the influence of pressure on the configurational stability. Fig.2 shows

the structure of the isomerase as obtained from a X-ray diffraction study of a single crystal [7].

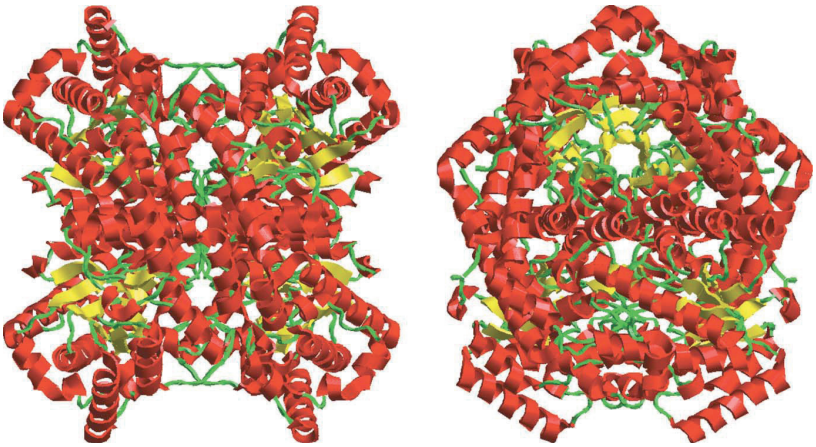


Fig. 2: *Structure of glucose/xylose isomerase from Streptomyces Rubiginosus. The pictures differ by a rotation clockwise by 90° around the y-axis.*

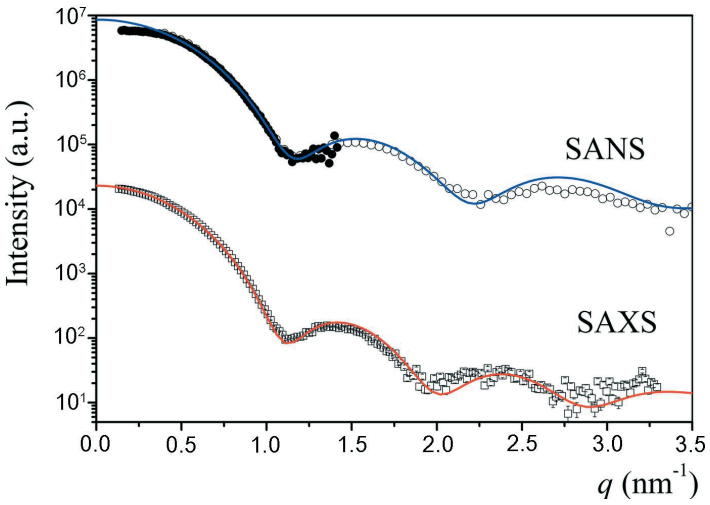


Fig. 3: *SANS and SAXS scattering curves (vertically shifted) in dilute solutions for the isomerase under conditions of a form factor measurement. Solid lines are fit to experimental curves using the CRYSON and CRYSOL simulation programs respectively, which stem on diffraction data of the isomerase.*

In Fig.3 we show the results from a small angle neutron and X-ray experiment from a dilute solution of the isomerase. The respectice scattering curves can be quite well described by a model (CRYSON and CRYSOL for SANS and SAXS data respectively) which takes the positions of the atoms into account as taken from the X-ray diffraction data of a single crystal. The models further make assumptions about the thickness of the hydration layer and the uniformity of the distribution of hydration water around the molecule. However, small deviations at high q point out the limits of these probably over-simpified approaches. The

deviations at low q for the SANS data are due to concentration effects, which can be accounted for by a proper modelling of the interparticle interactions.

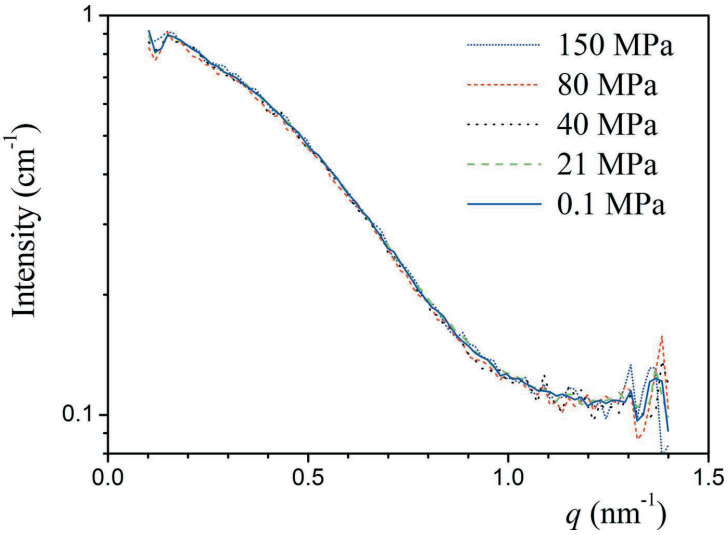


Fig. 4: *SANS scattering curves at various pressures as indicated in dilute solutions for the isomerase under conditions of a form factor measurement.*

In Fig.4 the influence of pressure on this structure is shown. The investigation clearly rules out that isomerase has a rather loose structure with a lot of voids, as the form factor at low q is invariant from pressure variations. Simple density effects, which would have changed the cross section have been taken into account preparing Fig.4. The isomerase molecule, from inspecting Fig.2, has, to a first approximation, a spherical shape and likely the active site is located in such a way that biochemical activity and pressure do not interfere much. This is usually different at much higher pressures, when the so called tertiary structure is altered and denaturation takes place. So it is a well-known fact that proteins do not change structure upon pressure treatment below the denaturizing pressure because likely the internal hydrogen bonding stabilizes a rather compact structure. This also explains unambiguously why in deepest sea fish and other creatures well survive and are seemingly not affected in their lives at a surrounding of 1kbar pressure. As an example for the transition to a denaturated state we show the change of diffusion behavior in lysozyme solutions as measured with dynamic light scattering as a function of pressure [8]. Here a pressure of more than 4.5kbar is needed to change the diffusion properties, which is indicative for a structural transformation. This is shown in Fig.5. Investigations of that kind are suited to exploit the stability limits between denaturated and native states of a protein as a function of pressure at a given temperature. Usually the application of high pressure results in the disruption of the native protein structure due to the decrease in the volume of the protein-solvent system upon unfolding. This volume change can be measured only by pressure changes as the variation of temperature would add on top also thermal energy. Thus a clear unbiased thermodynamic quantity, the $\Delta V_{\text{fold-unfold}}$ is accessible by pressure dependent measurements of the native to denaturated (unfolded) state of a protein. Changes in the tertiary structure induced by high pressures are usually reversible, because here only forming and releasing of non-covalent bonds are involved (no thermal heat load). This differs from the situation, where the system is subjected to high temperatures. In

these cases proteins, likewise if denaturatio is induced by chemicals, do not react elastically and reversibly.

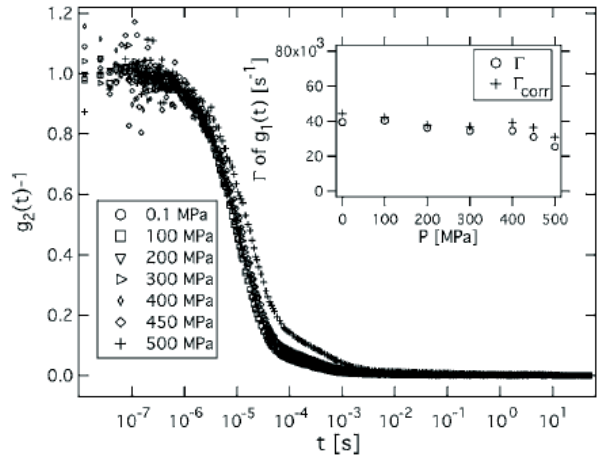


Fig. 5: *Normalized intensity autocorrelation function of a dilute buffered aqueous lysozyme solution as a function of pressure.*

However, also other scenarios are observable, where pressure has a larger effect on complex structures. We refer to the study of Winter *et al.* [9] on the pressure response of the protein Staphylococcal Nuclease , "Snase". In Fig.6 the SAXS scattering curve for Snase is shown. The solid line is a fit by CRYSOL as in Fig.3, based on X-ray diffraction data. What is evident from the Snase picture shown in Fig.6, is the relatively loose structure of the Snase as compared to the isomerase shown in Fig.2, which has a globular, more spherical, compact structure.

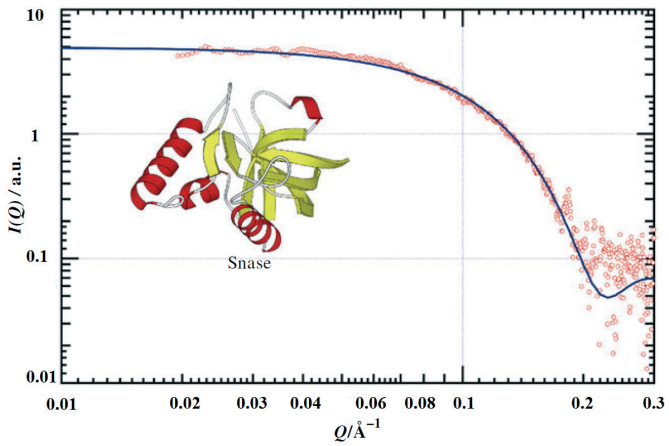


Fig. 6: *SAXS scattering curves for Snase in dilute solutions under conditions of a form factor measurement for P=1kbar.*

Varying the pressure gives at about a starting pressure of about 2kbar a change of structure, as shown in Fig.7, the pressure induced denaturation, which leads to a largely elongated structure, an ellipsoidal–like structure, as indicated in the lower part of Fig.7. Here the situation is reminiscent to the occurance of obviously large voids in the structure and hence pressure can rather easily induce structural changes at not too high pressures.

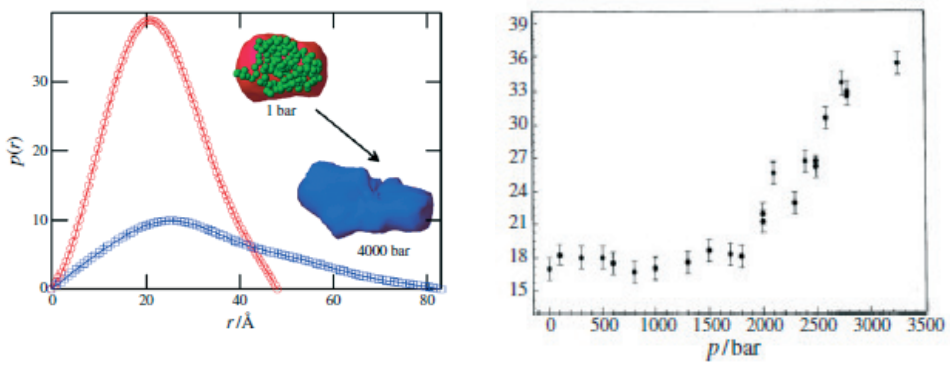


Fig. 7: *The pair-distance distribution function P(r) of Snase at 1bar and at 4kbar in the upper part of the Figure, in the lower part change of radius of gyration R_g/Å as a function of pressure.*

In Ref.[9] a rather extensive discussion can be found to what extent the pressure induced denaturation is contrasted to the temperature induced denaturation. Comparing the change of R_g as a function of P and T one finds additional unfolding processes in the latter case. Thus the interference effects due to thermal activation play an important role here and show that pressure is the relevant quantity to vary if entirely volume effects are considered.

2.2. Pressure Jump Techniques to Study Phase Behavior in Colloidal Sytems

It seems obvious that fast pressure variation so to say a jump is perfectly suited to study the kinetics of for example phase transitions. Temperature can also be used, however, depending on sample size and relaxation times involved, mainly the temperature profile of the process is detected and not the system's relevant process. Also a temperature change to a lower T is tricky as it is usually more difficult to take energy out of the system than rather to put it in. Here pressure jump techniques are superior. They do not suffer from the previously mentioned deficiencies. The time constants with which pressure can be applied on to a system can vary a bit and range from a 200bar/s with a hand operated hydraulic pump to 1kbar/s if that is automatized via a motor drive for the pump.

As an example for a nice application of the aformentioned pressure jump techniques the phase behavior of a sterically stabilized colloidal suspension is investigated [10]. The colloidal suspension consists of hairy silica spheres in toluene. This system shows phase separation into a silica rich phase and a silica poor phase (a liquid–gas phase transition) on varying temperature or pressure. On top of that also a percolation line intersects the phase transition. This percolation threshold depends on concentration and can be visualized as a gelation process which is kinetically controlled. In Fig.8 the situation is schematically shown.

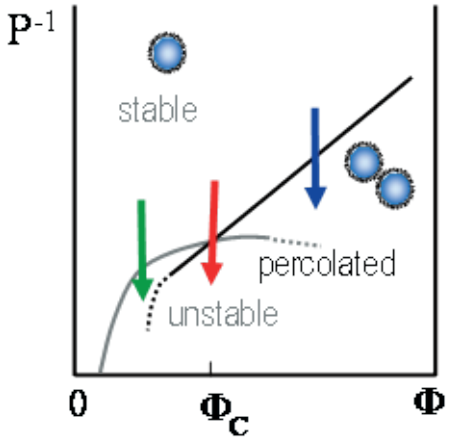


Fig. 8: *Schematic phase diagram of a sterically stabilized colloidal suspension. Φ_c is the critical concentration. The system goes from unstable (demixed state) by decreasing P (or likewise increasing temperature) to the stable state. Arrows indicate possible pressure jumps.*

As indicated in Fig.8, there were different presure jumps scenarios investigated which differ in the order where the various stability criteria were met. At low volume fraction the coexistence line was crossed prior to the percolation wheras at the highest volume fraction the percolation was crossed first upon increasing the pressure respectively. In order to get a hand on the different lines occuring in Fig.8, we have to find means to determine these by proper experiments. At the coexistence line usually demixing (pointing towards instability) takes place and hence any scattering technique is appropriate to detect it, as then the scattering intensity will strongly increase. The static structure factor will show a distinct q-behavior (q being the scattering vector) characteristic for nucleation and growth. In case of spinodal decomposition the static structure factor shows only the typical Ornstein-Zernicke scattering behavior. In our example here we used both light scattering and small angle neutron scattering to examine the coexistence line together with the spinodal. The percolation threshold, on the other hand, is accessible via a diffusive wave spectroscopy scattering experiment. It makes basically use from the fact that a percolated state is defined as being non-ergodic and hence the correlation functions of the diffusive wave spectroscopy scattered light will decay to a non-zero baseline. This behavior was studied for a the blue arrow sample as shown in Fig.8 and the result is shown in Fig.9.

From this kind of measurements as shown exemplarily in Fig.9 the percolation threshold can be determined. The question now was, what is the temporal evaluation of the system's scattering intensity extrapolated to vanishing q after a pressure jump and why is that an important question? The intensity I at q=0 is related to thermodynamics via $\partial\Pi/\partial\varphi \sim I(0)^{-1}$, where Π is the osmotic pressure and thus the phase equilibrium conditions are determined. Following the blue arrow from Fig.8, we have performed pressure jumps from ambient to the indicated pressure as shown in Fig.10.

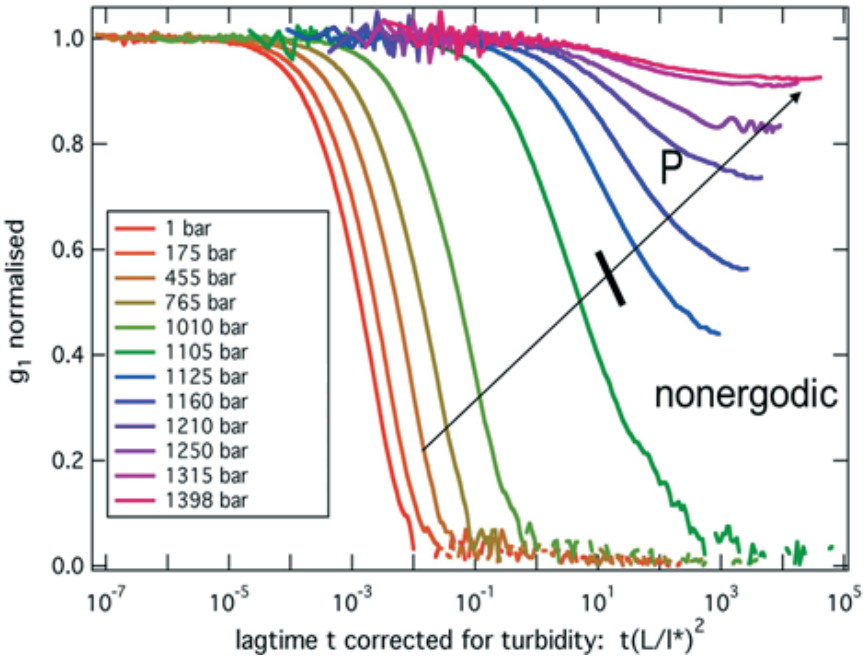


Fig. 9: *The normalized first order correlation function of the diffusive wave spectroscopy scattered light at different pressures. For P > 1105bar the system gets non-ergodic, thus defining the percolation threshold.*

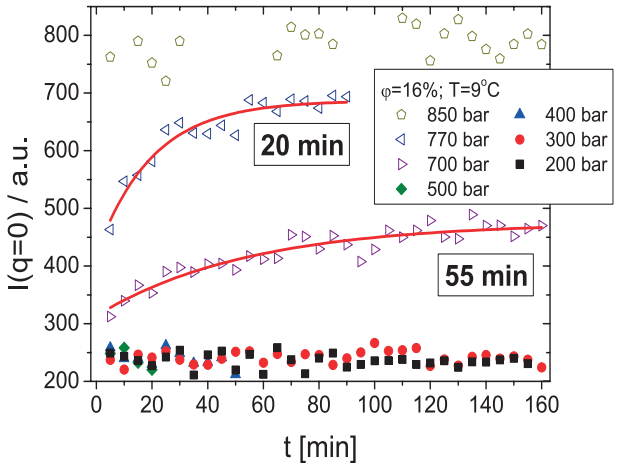


Fig. 10: *The forward scattering intensity I(0) as a function of time after pressure jumps from 1bar to the indicated pressures. For P≥700bar the system gets non-ergodic as determined by diffusive wave spectroscopy exemplarily shown in Fig.9. Time constants result from single exponential fits to the data.*

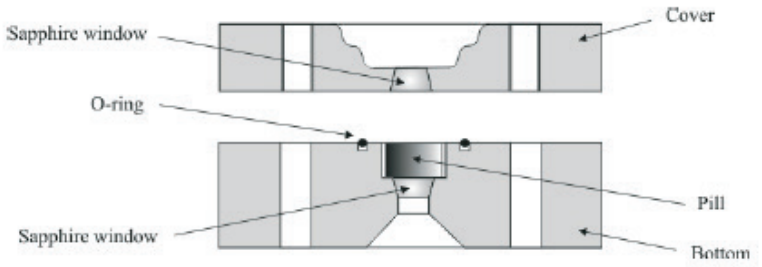
The intensity shows a time dependence after the percolation threshold was crossed. This is by no means obvious as usually a gelled or percolated, so to say kinetically frustrated state, is considered to be time invariant in its structure. This seems not to be the case and the reason for this is buried in the fact that the system's equilibrium properties are determined by thermodynamics and hence the magnitude of the forward intensity is governed by the proximity to the spinodal, which is reached, cf. Fig.8, at much higher pressures.

3 High Pressure Microscopy

3.1 Pressure Cells for Microscopes

Microscopy at high pressures is interesting for several reasons. High pressure microscopy for example is a versatile tool for investigating the microbial inactivation in favor over thermal treatment as pressure does not produce so many unwanted chemical reactions. *In situ* observations of this process are indispensible. This has been demonstrated by Frey et al. [1] who also give a rather complete overview of what is known in the field. Other promising fields of application are *in situ* observations of cellular systems and dynamic studies of reacting systems by spectroscopic investigations, where especially modern fluorescence techniques play an important role.

The combination of microscopy and high pressure is challenging, however, technically difficult since the main requirements for high resolution microscopy and for applying high pressure are in conflict. High mechanical resistance needed for high pressure automatically leads to an extended size of the sample environment (cell), while the working distance for the microscope objective needs to be minimal to assure maximum resolution.



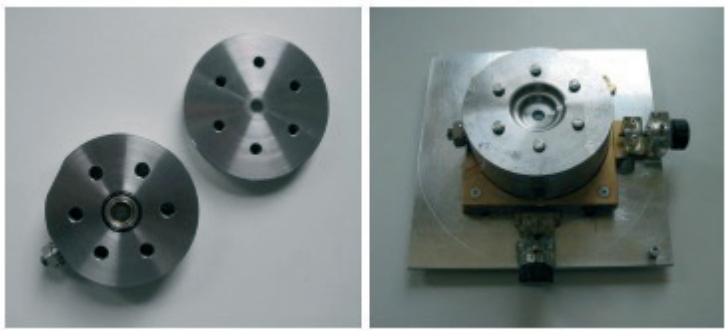


Fig. 11: *Upper part: Sketch of the high pressure microscopy cell. Lower part: Photographs of it. On the left dismantled, the brass pill is seen inside, and on the right, build in the microscopy stage to align in x-y directions. The pressure inlet is on the side of the cell. Pressure is generated via a hydraulic oil pump.*

The first working high pressure microscopy cell was developed by Hartmann et al. [11]. The cell works fine up to 300MPa producing decent microscopic pictures. However, the sample environment and change seems to be complicated and since a sapphire as optical window is used the birefringence properties are mediocre. The objective to sample distance is also quite large mainly due to the fact that the cell can withstand 300MPa pressure. The main reason for the blurred pictures as obtained in a recent application study [1], is however that the correction for thick windows is insufficient even when long working distance objectives are used.

Besides the up-to-date fluorescence techniques, it can in addition be important to preserve the plane of polarization. This is needed for pressure dependent measurements of the birefringence which is shown in chapter 2.2. In that research a small high pressure cell was employed using the idea of a separate sample container, the pill, [8], which is attached to two optical windows controlling both the stability and optical properties of the overall system. In Fig.11 the cell is depicted schematically. In the cover plate the space needed for the microscope objective to come as close as possible to the sample is indicated. Saphire is used as window material which is in principle optical birefringened. However, the orientation of the crystals were chosen such that the degree of birefrigence was minimized.

To overcome the specific deficiencies related to the use of saphire window to achieve thus superior optical properties concerning birefrigence glass windows should be used. However, since glass windows are less stable mechanically than sapphire or diamond, relatively thick windows must be used, so that no high magnification and hence no high resolution can be obtained. Another drawback is the separate sample container. It uses additional space through the windows of this pill and hence is also not suited for maintaining an as short as possible sample objective distance. Therefore a relatively simple cell for microscopic studies was developed which uses a thin diamond window as cover.

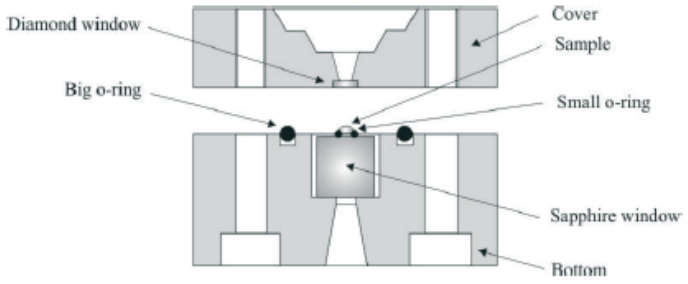


Fig. 12: *Sketch of a high pressure microscopy cell with easy sample handling.*

The top plate containing the diamond leaves enough space to come with the microscope objective as close as possible. The correction ring attached to the used long working distance objective is just sufficient to compensate for the large value of the diamonds index of

refraction. With this cell, shown schematically in Fig.12, 1kbar of pressure can be realized with a relatively easy sample handling.

3.2 Soft Matter Science Applications

Fd virus is a filamentous bacteriophage that can be considered as a colloidal rod. Its length is 880nm, the thickness 6.6nm and its persistence length is about 3µm. Water dispersions exhibit several degrees of ordering and, as biological species characterized by high mono-dispersity and easy production, can excellently serve as model liquid crystals. Phase transitions observed in *fd* are purely concentration dependent and, as a consequence, it means pressurizing can cause concentration changes followed by possible changes in ordering. The simple assumption is that the solvent has a different compressibility as *fd* has and a pressure change will thus cause a relative change in concentration, leading to an isotropic –nematic phase transition if the starting condition were chosen properly. How this transition looks under crossed polarizers is shown in Fig13. In the below described experiments now, pressure induced isotropic-nematic and nematic-isotropic phase transitions were observed. Moreover, high pressure was used to explore biphasic region of the phase diagram of *fd* to determine the limits of stability and instability by finding the loci of binodals and spinodals as a function of applied pressure. Experimental methods used included microscopy, optical birefringence and transmission measurements.

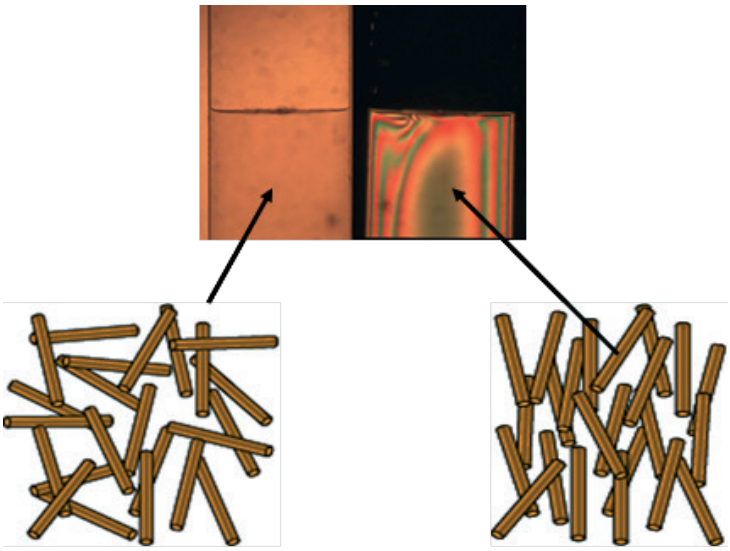


Fig. 13: *Sketch of fd suspensions: lower left indicates isotropic phase and leads to homogenous picture in the transmission microscope, upper panel, under conditions of crossed polarizers, whereas at lower right, a nematic ordering of the fd is present, leading to optical birefringence seen under crossed polarizers.*

Such an experiment has been undertaken, where a custom made high pressure microscopy cell was used [12]. The experiment was performed such that the pressure was increased with a rate of about 100bar/sec to reach the starting pressure and the resulting microscopy pictures in transmission under crossed polarizers were detected as a function of time. The result is shown in Fig.14.

Whereas for the lower starting pressures 200 and 400bars no transition can be observed, areas with higher orientation (higher order parameter) appear at higher starting pressures. The observed time behavior differs in the magnitude of the induction time after which structured or ordered regions occur. At 1kbar the almost full ordering is achieved instantanously indicating spinodal decomposition, wheras at lower starting pressures a noticable induction time is seen, indicating a nucleation and growth mechanism (binodal decomposition).

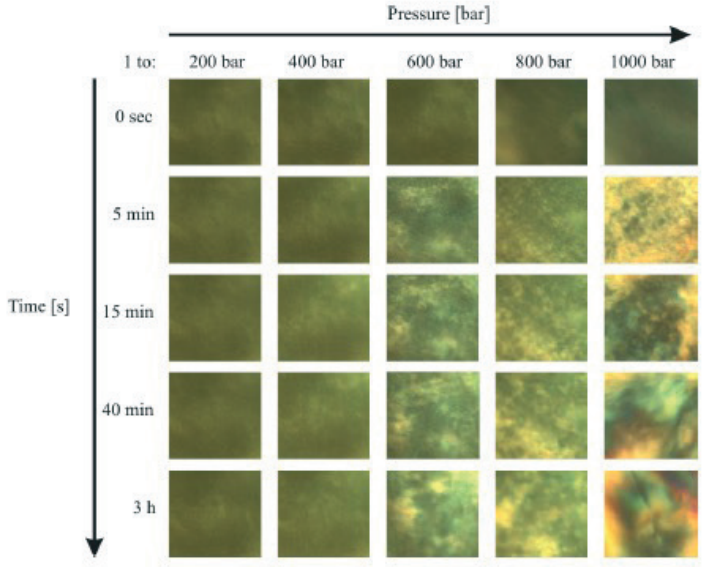


Fig. 14: *The effect of pressure on the isotropic-nematic phase transition as a function of waiting time for different starting pressures after a jump from 1bar for an aqueous solution of fd.*

This scenario is also possible in the reversed case, means we start from a high pressure and release it with a rate of 100bar/s. This and the analysis of the data from Fig.14 are used to construct a phase diagram, which is very complicated to obtain otherwise because every concentration must be prepared separately.

References

[1] B. Frey, M. Hartmann, M. Herrmann, R. Meyer-Pittroff, K. Sommer and G. Bluemelhuber, Microscopy Research and Technology 69, 65 (2006)
[2] P. Baaske, F.M. Weinert, S. Duhr, K.H. Lemke, M.J. Russell, and D. Braun, P. Natl. Acad. Sci. USA, 104, 9346 (2007)
[3] http://iodp.tamu.edu/publications/proceedings.html
[4] B. Bhushan, Introduction to Tribology (Wiley, New-York, 2002)
[5] A. Drljaca, C.D. Hubbard, R. van Eldik, T. Asano, M.V. Basilevsky and W.J. le Noble, Activation and Reaction Volumes in Solution, Chemical Reviews 98 (6), 2167 (1998)
[6] Frontiers in High Pressure Biochemistry and Biophysics, Biochemica et Biophysica Acta, 1595, 1-403 (2002)
[7] E. Banachowicz, M. Kozak, A. Patkowski, G. Meier and J. Kohlbrecher, J Appl Cryst. 42, 461 (2009)
[8] J. Kohlbrecher, A. Bollhalder, R. Vavrin and G. Meier, Rev. Sci. Instrum. 78, 125101 (2007)
[9] R. Winter and R. Köhling, J. Phys: Condens. Matter 16, S327 (2004)
[10] J. Kohlbrecher, J. Buitenhuis, G. Meier and M.P. Lettinga, J. Chem. Phys. 125, 044715 (2006); R. Vavrin, J. Kohlbrecher, A. Wilk, M. Ratajczyk, M.P. Lettinga, J. Buitenhuis and G. Meier, J. Chem. Phys. 130, 154903 (2009)
[11] M. Hartmann, F. Pfeifer, G. Dornheim and K. Sommer, HPDS-Hochdruckzelle zur Beobachtung mikroskopischer Phänomene unter Hochdruck, Chemie Ingenieur Technik 75 1763 (2003)
[12] P. Holmqvist, M. Ratajczyk, G. Meier, H.H. Wensink and M.P. Lettinga, Phys. Rev. E, 80, 031402 (2009)

F 3 Charged Colloids and Electric Fields

Kyongok Kang and Jan K.G. Dhont
Institute of Complex Systems / ICS-3: Soft Condensed Matter
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 The Standard Electro-Kinetic Equations **3**

3 The Poisson-Boltzmann Equation and Debye-Hückel Approximation **5**

4 The Electric Double Layer and Ion Condensation **6**
4.1 The electric double layer around a spherical colloid 6
4.2 The double layer at a flat interface, and ion-condensation 7

5 Double-Layer Interactions of Flat Interfaces **10**
5.1 The free energy of an electric double layer 11
5.2 The interaction potential 13

6 Electrode polarization **15**

7 Electrophoretic Mobility **19**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

In this introductory lecture the interest is in some of the basic properties of charged colloids, also in the presence of an external electric field. The charge of macromolecules, including for example synthetic colloids, proteins, DNA and virus particles, originates from charged groups (like carboxyl and amino groups) that are chemically attached to the surface of the macromolecule. In this lecture we discuss mostly spherical colloids. The colloid is embedded in a solvent that contains solvated ions which interact through Coulombic forces with the charged colloidal surface. There is a thin region at the surface, where structural order varies on length scales comparable to the size of the ions. This layer is referred to as the *Stern layer*, and is depicted in Fig.1. It accounts for the finite size of potential determining groups that are chemically attached to the surface, as well as possible adsorbed ions. For dynamic processes, like in electrophoresis, the relatively small mobility of ions and water molecules close to the surface is of significance. These mobilities near the surface are much smaller than those in bulk, away from the surface. The region within which the mobilities are relatively small, usually extends beyond the above mentioned *static Stern layer*, and is referred to as the *dynamic Stern layer*. In electrophoresis, for example, a *shear-plane* or *slipping-plane* is defined, at which there are effective stick boundary conditions for water flow. This shear-plane is located outside the static Stern layer, and defines the extent of the dynamic Stern layer. Due to the thermal motion of ions in solution, there is a charge distribution of finite extent around the macromolecule, which is referred to as the *electric double layer*. The extent of the double layer is typically much larger than the size of the ions (see Fig.1).

The simplest realistic model for the electric double layer is *the Gouy-Chapman model* (~ 1910), which assumes a perfectly smooth colloidal surface, and ions to be represented by point charges. This model takes thermal diffusion of the ions into account, but disregards the detailed structure of the surface. Accounting for finite-size effects requires quite complex theoretical approaches, while the Stern layer is specific to the chemistry of the surface of the colloidal particle and the ions present in solution, which is both beyond the scope of this introductory lecture. The lecture is thus restricted to Gouy-Chapman like approaches, where the details of the surface of the colloids are disregarded.

The equations of motion that are at the basis of the Gouy-Chapman theory will be derived in section 2. These so-called standard electro-kinetic equations are at the basis of a large body of literature concerned with the analysis of properties of colloids, also in external fields. In equilibrium, these standard electro-kinetic equations reduce to the so-called Poisson-Boltzmann equation which describes the structure of the double layer. The Poisson-Boltzmann equation and its linearized version is discussed in section 3. Linearization of equations of motion, and the Poisson-Boltzmann equation in particular, with respect to the surface potential/charge of the colloid is commonly referred to as the Debye-Hückel approximation. The double-layer structure is analyzed in section 4 on the basis of the Poisson-Boltzmann equation, including a brief discussion of ion-condensation. The interaction between charged colloids as their double layers overlap is discussed in section 5. An important ingredient in the calculation of the interaction potential is the Helmholtz free energy of double layers, which will be discussed in some detail. When an electric field is applied to a suspension of charged colloids, the development of electric double layers at the electcrodes diminishes the field strength in bulk, away from the electrodes. This phenomenon is referred to as electrode polarization. As the colloids experience the attenuated field, it is important to know how the screening of the electric field depends on the frequency of the applied field, which is the topic of section 6. Finally, section 7 deals

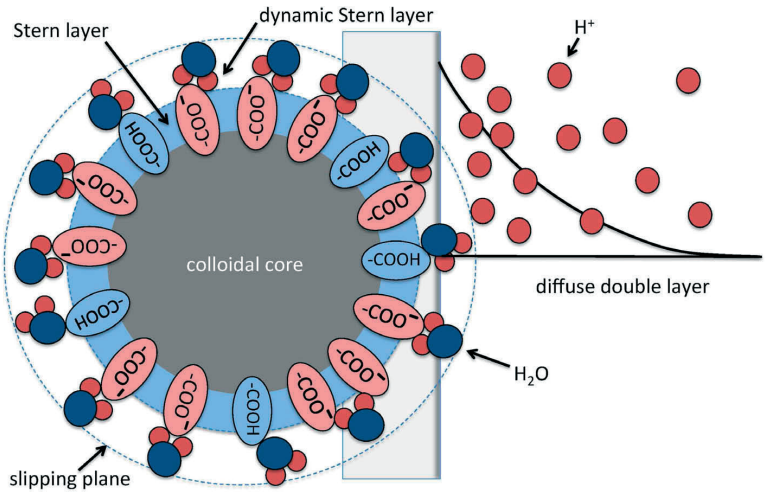


Fig. 1: *A sketch of the (static) Stern layer, indicated in blue, which contains in this example carboxyl groups that are chemically bounded to the colloidal core. The dynamic Stern layer includes a solvation layer consisting of water molecules associated to the core material of the colloid as well as the carboxyl groups. The dynamic Stern layer is bounded by the slipping layer. The curve on the right shows the concentration of H^+-ions, dissociated from the carvoxyl groups, within the diffuse electric double layer as a function of the distance from the colloidal surface.*

with electrophoresis, that is, the velocity that a charged colloid attains under the influence of an external field, which can be used to determine the surface charge/potential of colloids.
A few of the books that are dedicated to charged colloids are those of Verwey and Overbeek [1], of Hunter [2], of Israelachvili [3], and Lyklema [4].

2 The Standard Electro-Kinetic Equations

We discuss here the so-called standard electro-kinetic equations, that constitute the basis of what is discussed in this lecture. As mentioned in the introduction, the theory discussed here is limited to smooth, unstructured surfaces and point-like ions. The aim of this section is to derive equations of motion for the concentration of ions within the diffuse electric double layer in the presence of a charged surface.
Let ρ_α denote the number-concentration (= number of particles per unit volume) of a certain ion species α with a charge $e\,z_\alpha$, with $e > 0$ the elementary charge, and where z_α is referred to as *the valency* of the ion species. When an ion is given an initial velocity, the thermal averaged velocity will decrease in time due to friction with the surrounding solvent molecules. The decay time of the initial velocity is referred to as the velocity-relaxation time. On a time scale that is much larger than this velocity-relaxation time, there is force balance, since the inertial forces can be neglected. This time scale is referred to as *the diffusive time scale*, or *the Brownian time scale*. That is, all (non-inertial) forces add up to zero. As an ion moves with an instantaneous velocity $\mathbf{v}$, the friction force is in the opposite direction, and is equal to $\mathbf{F}^{fric} = -\zeta\,\mathbf{v}$, where ζ is the friction coefficient. The fluid may be in motion, in which case the velocity should

be taken relative to the local fluid flow velocity $\mathbf{u}$, so that $\mathbf{F}^{fric} = -\zeta\,(\,\mathbf{v}-\mathbf{u}\,)$. There is an electrostatic force equal to $\mathbf{F}^{E} = -e\,z_\alpha \nabla\Psi$, where ∇ is the gradient operator with respect to the position coordinate of the ion, and Ψ is the local electrostatic potential. This potential is due to the presence of the charged interface as well as the resulting inhomogeneous charge distribution within the electric double layer. A third force that is always present on the diffusive time scale is the Brownian force $\mathbf{F}^{Br} = -\,k_BT\nabla\ln\rho_\alpha$ on the ion, where k_B is Boltzmann's constant, and T is the absolute temperature (for more details on the Brownian force and the diffusive time scale, see **chapter B1**). Force balance implies that $\mathbf{F}^{fric}+\mathbf{F}^{E}+\mathbf{F}^{Br}=\mathbf{0}$. It thus follows that the thermally averaged velocity is equal to,

$$\mathbf{v} \;=\; -\,D\,\{\,\nabla\,\ln\rho_\alpha + \beta\,e\,z_\alpha\,\nabla\Psi\,\} + \mathbf{u}\;,$$

where $\beta = 1/k_BT$, and,

$$D \;=\; \frac{k_BT}{\zeta}\;,$$

is the ion's diffusion coefficient. The flux of ions of species α is thus,

$$\mathbf{j}_\alpha \;=\; \rho_\alpha\,\mathbf{v} \;=\; -\,D\,\{\,\nabla\,\rho_\alpha + \beta\,e\,z_\alpha\,\rho_\alpha\,\nabla\Psi\,\} + \rho_\alpha\,\mathbf{u}\;. \tag{1}$$

Substitution of this expression into the continuity equation (see **chapter B1** for a derivation of the continuity equation),

$$\frac{\partial\rho_\alpha}{\partial\,t} \;=\; -\,\nabla\cdot\mathbf{j}_\alpha\;,$$

leads to the standard-electro-kinetic equation of motion for the number concentration of ions,

$$\frac{\partial\,\rho_\alpha(\mathbf{r},t)}{\partial\,t} \;=\; D\,\nabla\cdot\{\,\nabla\rho_\alpha(\mathbf{r},t) + \beta\,e\,z_\alpha\,\rho_\alpha(\mathbf{r},t)\,\nabla\Psi(\mathbf{r},t)\,\} - \nabla\cdot\{\,\rho_\alpha(\mathbf{r},t)\,\mathbf{u}(\mathbf{r},t)\,\}\;, \tag{2}$$

where now the dependencies on the position $\mathbf{r}$ and time t are denoted explicitly.
These equations of motion for ion-concentrations couple to the electrostatic potential Ψ. For the slow processes under consideration, the potential essentially adjusts instantaneously to the charge density. The connection between the potential and charge density is thus given by the Poisson equation from electrostatics,

$$\nabla^2\Psi(\mathbf{r},t) \;=\; -\,\frac{\rho(\mathbf{r},t)}{\epsilon_s}\;, \tag{3}$$

where ϵ_s is the dielectric constant of the solvent (which accounts for the diminished charges due to polarization of the solvent), and where the charge density within the solvent is by definition related to the ion-concentrations as,

$$\rho(\mathbf{r},t) \;=\; \sum_\alpha e\,z_\alpha\,\rho_\alpha(\mathbf{r},t)\;. \tag{4}$$

The equations of motion (2,3,4) are the standard electro-kinetic equations.
In principle, additional equations of motion must be derived that describe the flow velocity $\mathbf{u}$. In this lecture the flow velocity is only of importance in section 7 on the electrophoretic mobility of colloids, and will be discussed there separately.

The standard electro-kinetic equations are supplemented by boundary conditions. Let $\mathcal{S}$ denote a surface that is not permeable for the ions (like the surface of a colloidal particle or electrodes). The component of the flux in eq.(1) normal to the surface thus vanishes,

$$\hat{\mathbf{n}}(\mathbf{r}) \cdot \{ \nabla \rho_\alpha(\mathbf{r},t) + \beta\, e\, z_i\, \rho_\alpha\, \nabla\Psi(\mathbf{r},t) \} \;=\; 0 \quad , \quad \text{for} \quad \mathbf{r} \in \mathcal{S} \; , \tag{5}$$

where the solvent flow is assumed to be zero. Here, $\hat{\mathbf{n}}$ is the unit vector (with length unity as indicated by the hat), which is locally perpendicular to the surface $\mathcal{S}$.
The above three equations (2-5) are at the basis of the Gouy-Chapman theory. As mentioned before, it assumes point-like ions, where excluded-volume interactions between the ions are absent, and regards the surface $\mathcal{S}$ as a mathematically smooth surface. In addition, this theory is a *mean-field approach*. The actual, fluctuating Coulomb force on an ion i is equal to,

$$\mathbf{F}^E \;=\; -\nabla_i \sum_{j\neq i} \frac{e^2\, z_i\, z_j}{4\,\pi\,\epsilon_s\;|\mathbf{r}_i - \mathbf{r}_j|} \; ,$$

where ∇_i is gradient operator with respect to the position coordinate $\mathbf{r}_i$ of ion i, and the sum is over all ions except for ion i. In the present approach, instead of using this microscopic expression for the electric force, the ion charges are "smeared out" to form a continuous charge distribution, giving rise to a potential with the accompanied force on ion i equal to $\mathbf{F}^E = -e\, z_\alpha\, \nabla\Psi$.
References [5, 6, 7, 8, 9, 10, 11] aim at the improvement of the Gouy-Chapman approach (which is just a very limited choice out of the vast body of existing literature on this subject).

3 The Poisson-Boltzmann Equation and Debye-Hückel Approximation

In equilibrium, in the absence of an external electric field, the fluid flow velocity is zero and the ion-concentrations and potential are time-independent, so that the equation of motion (2) reduces to,

$$\begin{aligned} \nabla\rho_\alpha(\mathbf{r}) \;=\; -\beta\, e\, z_\alpha\, \rho_\alpha(\mathbf{r})\, \nabla\Psi(\mathbf{r}) \quad &\rightarrow \quad \nabla \ln \rho_\alpha(\mathbf{r}) \;=\; -\,\beta\, e\, z_\alpha\, \nabla\Psi(\mathbf{r}) \\ &\rightarrow \quad \rho_\alpha(\mathbf{r}) \sim \exp\{-\beta\, e\, z_\alpha\, \Psi(\mathbf{r})\} \; . \end{aligned}$$

The proportionality constant is the concentration ρ_α^0 outside the double layer, where the potential $\Psi = 0$. Hence,

$$\rho_\alpha(\mathbf{r}) \;=\; \rho_\alpha^0\, \exp\{-\beta\, e\, z_\alpha\, \Psi(\mathbf{r})\} \; . \tag{6}$$

Substitution into eq.(4) for the charge density, and subsequent substitution into the Poisson equation (3), it is found that,

$$\nabla^2\Psi(\mathbf{r}) \;=\; -\sum_\alpha \frac{e\, z_\alpha}{\epsilon_s}\, \rho_\alpha^0\, \exp\{-\beta\, e\, z_\alpha\, \Psi(\mathbf{r})\} \; . \tag{7}$$

This is a closed equation for the potential, which is referred to as *Poisson-Boltzmann equation*.

For small potentials, such that $\mid e\, z_\alpha\, \Psi \mid < 1$ (which means that the potential energy is small compared to the thermal energy k_BT), the exponential function can be linearized, using that $\exp\{x\} = 1 + x + \frac{1}{2}\, x^2 + \cdots$, leading to,

$$\nabla^2 \Psi(\mathbf{r}) \;=\; -\sum_\alpha \frac{e\, z_\alpha}{\epsilon_s}\, \rho_\alpha^0\, [\, 1 - \beta\, e\, z_\alpha\, \Psi(\mathbf{r})\,] \;.$$

Such a linaerization is commonly referred to as *the Debye-Hückel approximation*. Since the solvent outside the double layer is uncharged, we have $\sum_\alpha e\, z_\alpha\, \rho_\alpha^0 = 0$, so that,

$$\nabla^2 \Psi(\mathbf{r}) \;=\; \kappa^2\, \Psi(\mathbf{r}) \;, \tag{8}$$

where the (inverse) *Debye length* κ is equal to as,

$$\kappa \;=\; \sqrt{\frac{e^2\, \beta}{\epsilon_s} \sum_\alpha z_\alpha^2\, \rho_\alpha^0} \;. \tag{9}$$

As will become clear later, the Debye length κ^{-1} measures the spatial extent of the diffuse electric double layer. According to the above result, the Debye length decreases with increasing ion-concentrations. The physical picture is that more ions screen the charge of the colloid more effectively. Equation (8) is referred to as *the linearized Poisson-Boltzmann equation*. The electric double layer theory based on the linearized Poisson-Boltzmann equation is also referred to as *the Debye-Hückel theory*.

4 The Electric Double Layer and Ion Condensation

The structure of the electrical double layer will be discussed here for a spherical colloidal particle within the linearized Poisson-Boltzmann approach, and for flat double layers with the inclusion of non-linear contributions. The double layer structure for such a flat interface is relevant for colloids with a very thin double layer, and for the analysis of electrode polarization in section 6. This analytical solution of the non-linear Poisson-Boltzmann equation for flat interfaces shows that at high potentials there is a strong accumulation of ions close to the surface, which is referred to as *ion-condensation*.

4.1 The electric double layer around a spherical colloid

For a spherical colloid, the Laplace operator is most conveniently expressed in spherical coordinates, where the angular part can be neglected as the double layer is spherically symmetric. We thus find that,

$$\frac{1}{r^2}\, \frac{d}{dr} \left(r^2\, \frac{d\, \Psi(r)}{dr} \right) \;=\; \kappa^2\, \Psi(r) \;,$$

where r is the distance from the center of the colloidal sphere. The general solution of this differential equation is (this solution can be found by introducing the auxiliary function $f(r) = r\, \Psi(r)$),

$$\Psi(r) \;=\; A\, \frac{\exp\{-\kappa\, r\}}{r} + B\, \frac{\exp\{+\kappa\, r\}}{r} \;,$$

where A and B are integration constants. Since for large distance r the potential vanishes, $B = 0$. When $r = a$ (the radius of the sphere), the potential is equal to the its value Ψ_s at the surface, so that $A = a\,\exp\{-\kappa\,a\}$, and hence,

$$\Psi(r) \;=\; \Psi_s\,\frac{\exp\{-\kappa\,(r-a)\}}{r/a}\;. \tag{10}$$

The charge density ρ is now obtained from the Poisson equation (3) together with eq.(8),

$$\rho(r) \;=\; -\epsilon_s\,\kappa^2\,\Psi_s\,\frac{\exp\{-\kappa\,(r-a)\}}{r/a}\;.$$

The surface potential can be related to the total charge Q of the colloid, by noting that Q is equal to minus the total charge residing within the double layer,

$$Q \;=\; -\int_{r>a} d\mathbf{r}\;\rho(r) \;=\; -\,4\,\pi\int_a^\infty dr\;r^2\,\rho(r) \;=\; 4\,\pi\,\epsilon_s\,a\,\Psi_s\;(\,1+\kappa\,a\,)\;, \tag{11}$$

so that the charge density can also be written as,

$$\rho(r) \;=\; -\,\frac{Q\,\kappa^2}{4\,\pi\,a\,(1+\kappa\,a)}\,\frac{\exp\{-\kappa\,(r-a)\}}{r/a}\;. \tag{12}$$

Note that it follows from eqs.(10,11) that when the salt concentration vanishes, so that $\kappa = 0$, the potential is equal to $\Psi(r) = Q/4\pi\epsilon_s r$, which is precisely Coulomb's law. From the above analysis it is evident that the Debye length κ^{-1} measures the extent of the electric double layer.

4.2 The double layer at a flat interface, and ion-condensation

Next consider a flat, uniformly charged interface. Let z denote the distance from the interface. The linearized Poisson-Boltzmann equation (8) now reads,

$$\frac{d^2}{d\,z^2}\,\Psi(z) \;=\; \kappa^2\,\Psi(z)\;,$$

the solution of which is a simple exponential decaying potential,

$$\Psi(z) \;=\; \Psi_s\,\exp\{-\,\kappa\,z\}\;.$$

Just as for the spherical colloid in the previous subsection, the charge density within the double layer is found to be equal to,

$$\rho(z) \;=\; -\epsilon_s\,\frac{d^2\Psi(z)}{d\,z^2} \;=\; -\,\epsilon_s\,\kappa^2\,\Psi(z) \;=\; -\,\epsilon\,\kappa^2\,\Psi_s\,\exp\{\,-\,\kappa\,z\,\}\;, \tag{13}$$

while the connection between the surface potential Ψ_s and the surface charge density σ on the plate is given by,

$$\sigma \;=\; -\int_0^\infty dz\;\rho(z) \;=\; \epsilon\,\kappa\,\Psi_s\;. \tag{14}$$

For the flat interface, the non-linear Poisson-Boltzmann equation also allows for an analytical solution for a symmetric electrolyte, for which we will take $z_\alpha = \pm 1$. Introducing the dimensionless potential and dimensionless distance,

$$\tilde{\Psi} \;=\; \beta\, e\, \Psi \quad , \quad Z = \kappa\, z \;, \tag{15}$$

the non-linear Poisson-Boltzmann equation can then be written as (where the "hyberbolic-sine function" is defined as $\sinh\{x\} = (\exp\{+x\} - \exp\{-x\})/2$),

$$\frac{d^2}{d\,Z^2}\,\tilde{\Psi}(Z) \;=\; \sinh\{\tilde{\Psi}(Z)\}\;. \tag{16}$$

A first integration can be done by multiplying both sides with $d\Psi/dZ$, and using that (with the "hyperbolic-cosine function" is defined as $\cosh\{x\} = (\exp\{+x\} + \exp\{-x\})/2$),

$$\begin{aligned}
\frac{d\tilde{\Psi}(Z)}{d\,Z}\,\frac{d^2\tilde{\Psi}(Z)}{d\,Z^2} \;&=\; \frac{1}{2}\,\frac{d}{d\,Z}\left(\frac{d\tilde{\Psi}(Z)}{d\,Z}\right)^2 \;, \\
\frac{d\tilde{\Psi}(Z)}{d\,Z}\,\sinh\,\tilde{\Psi}(Z) \;&=\; \frac{d}{d\,Z}\,\cosh\{\tilde{\Psi}(Z)\}\;.
\end{aligned}$$

Since the potential vanishes for large Z, we thus find that,

$$\frac{d}{d\,Z}\,\tilde{\Psi}(Z) \;=\; \sqrt{2\,\Big(\cosh\{\tilde{\Psi}(Z)\} - 1\Big)} \;=\; -2\,\sinh\{\tilde{\Psi}(Z)/2\}\;. \tag{17}$$

The second integration can be done as follows. First define $\varphi(Z) = \tanh\{\tilde{\Psi}(Z)/4\}$ (where the "hyperbolic-tangent function" is defined as $\tanh x = \sinh x/\cosh x$). From the above result for the first-order derivative, one finds that $d\varphi(Z)/dZ = -\,\varphi(Z)$, and hence $\varphi(Z) = \varphi_s\,\exp\{-Z\}$, where $\varphi_s = \tanh\{\tilde{\Psi}_s/4\}$ with $\tilde{\Psi}_s$ the value of the dimensionless potential at the surface. We thus have,

$$\tanh\{\tilde{\Psi}(Z)/4\} \;=\; \tanh\{\tilde{\Psi}_s/4\}\,\exp\{-Z\}\;. \tag{18}$$

It follows from this relation that, [1]

$$\tilde{\Psi}(Z) \;=\; 2\,\ln\left\{\frac{1 + \tanh\{\tilde{\Psi}_s/4\}\,\exp\{-Z\}}{1 - \tanh\{\tilde{\Psi}_s/4\}\,\exp\{-Z\}}\right\}\;. \tag{20}$$

[1] This inversion is done as follows. Let $\alpha = \exp\{\Psi(Z)/4\}$, and $\xi = \tanh\{\Psi_s/4\}\,\exp\{-Z\}$, then eq.(18) reads,

$$\tanh\{\tilde{\Psi}(Z)/4\} \;=\; \frac{\exp\{\tilde{\Psi}(Z)/4\} - \exp\{-\tilde{\Psi}(Z)/4\}}{\exp\{\tilde{\Psi}(Z)/4\} + \exp\{-\tilde{\Psi}(Z)/4\}} \;=\; \frac{\alpha - \frac{1}{\alpha}}{\alpha + \frac{1}{\alpha}} \;=\; \xi\;.$$

This is easily rewritten as,

$$\alpha^2 - 1 \;=\; \xi\,\left(\alpha^2 + 1\right) \quad \rightarrow \quad \alpha \;=\; \sqrt{\frac{1+\xi}{1-\xi}}\;. \tag{19}$$

Substitution of the original variables immediately leads to eq.(20).

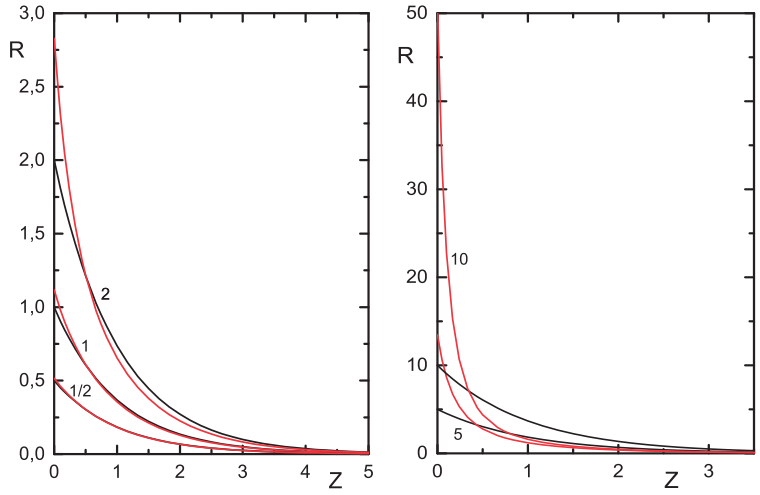


Fig. 2: *The dimensionless charge density $R = \rho(Z)/(-2\,e\,\rho^0)$ as a function of the dimensionless distance $Z = \kappa\,z$, for various values of $\tilde{\gamma} = \beta\,e\,\sigma/(\epsilon_s\,\kappa)$, as indicated in the plots. The black lines are for the linear theory, the red lines for the non-linear theory.*

The relation between the surface charge density and the derivative of the potential at the surface follows from Poisson's equation (3),

$$\frac{d\,\tilde{\Psi}(z)}{d\,z}_{|z=0} = -\frac{\sigma}{\epsilon_s} \quad \rightarrow \quad \frac{d\,\tilde{\Psi}(Z)}{d\,Z}_{|Z=0} = -\frac{\beta\,e\,\sigma}{\kappa\,\epsilon_s} = -2\,\sinh\{\tilde{\Psi}_s/2\}\;,$$

where in the last equation we used eq.(17) for the first-order derivative. A similar inversion as before thus leads to,

$$\tilde{\Psi}_s(\sigma) = 2\,\ln\left\{\frac{\beta\,e\,\sigma}{2\,\kappa\,\epsilon_s} + \sqrt{1+\left(\frac{\beta\,e\,\sigma}{2\,\kappa\,\epsilon_s}\right)^2}\right\}\;, \tag{21}$$

which is the connection between the surface potential and the surface charge density. Upon linearization with respect to σ, this reproduces the result in eq.(14) that was obtained from the linearized Poisson-Boltzmann equation, as it should (note that here Ψ_s is the dimensionless potential in eq.(15)). The charge density in the double layer follows from the Poisson equation, eq.(16), and eq.(9) (which for the present case reads $\kappa^2 = 2\,\beta\,e^2\,\rho^0/\epsilon_s$, with ρ^0 the number concentration of single ion species outside the double layer),

$$\rho(Z) = -\,2\,e\,\rho^0\,\sinh\left[2\,\ln\left\{\frac{1+\tanh\{\tilde{\Psi}_s(\sigma)/4\}\,\exp\{-Z\}}{1-\tanh\{\tilde{\Psi}_s(\sigma)/4\}\,\exp\{-Z\}}\right\}\right]\;. \tag{22}$$

In Fig.2 the quantity $R = \rho(Z)/(-2\,e\,\rho^0)$ is plotted as a function of Z, for various values of $\tilde{\gamma} = \beta\,e\,\sigma/(\epsilon_s\,\kappa)$ which is a dimensionless surface charge density. The dimensionless number $\tilde{\gamma}$

measures the degree of non-linearity. For $\tilde{\gamma} < 1$, the linear and non-linear theory are not very different. Strong deviations are evident for larger values of $\tilde{\gamma}$, as can be seen from the plots on the right in Fig.2.
The very sharp increase of the ion density close to the wall for larger values of the surface charge density is referred to as "ion-condensation". There is a thin layer of very high ion concentration close to the wall (especially clear for $\tilde{\gamma} = 10$ in Fig.2), which screens the surface charge density σ. Further away from the wall, where the potential is small, the non-linear theory reduces to the linear theory. In order to intuitively understand ion-condensation, we need to introduce the so-called *Bjerrum length* l_B. This is the length over which two elementary charges have a Coulomb energy equal to the average kinetic energy k_BT of ions in solution. By definition $k_BT = e^2/(4\pi\epsilon_s l_B)$, and hence,

$$l_B \;=\; \frac{\beta\, e^2}{4\,\pi\,\epsilon_s}\;.$$

For water at room temperature $l_B = 0.74 \times 10^{-9}\, m$. It is easily seen that $\tilde{\gamma} = 4\pi\,(l_B/b)\,(\kappa^{-1}/b)$. This result has the following intuitive interpretation, which is at the basis of classical ion-condensation theory. When two chemically bound surface charges are separated by a distance less than l_B, it is energetically favorable that these groups are neutralized through the association of ions in solution: the ions lose their kinetic energy, but more energy is gained by eliminating the Coulombic interaction between the two bounded charges. More precisely, the number of ions interacting with a given bounded ion is of the order $\kappa^{-2}\,(\sigma/e)$, since interactions of ions separated by a distance larger than κ^{-1} are screened. Since the surface charge density σ is of the order e/b^2, with b the distance between chemically attached surface groups, the interaction energy of a bounded ion with remaining bounded ions is of the order $(e^2/4\pi\epsilon_s\kappa^{-1}) \times (\kappa^{-2}/b^2)$. The interaction energy in units of the thermal energy k_BT is thus of the order $(l_B/b)\,(\kappa^{-1}/b)$. Apart from a prefactor 4π, this is precisely the expression given above for $\tilde{\gamma}$.
Ion-condensation is first described by Manning, and is therefore commonly referred to as *Manning condensation*. More information on ion-condesation theory can be found in Refs.[12, 13, 14] which, again, is just a very biased choice from the vast body of existing lirerature on this subject.

5 Double-Layer Interactions of Flat Interfaces

Since the colloids are very large compared to the ions, and therefore diffuse much slower than the ions, the double layers may be considered as being in equilibrium with the instantaneous configuration of the colloids. Poisson-Boltzmann theory can thus be employed to calculate the charge distribution within the double layer in the presence of two colloids with fixed positions. In subsection 5.1, a general expression for the Helmholtz free energy of a double layer is derived. An important point here is that, for the commonly adopted simple adsorption equation of state of the potential determining ions, there can only be (instantaneous) equilibrium when the surface potential of the two colloids is constant, independent of the distance between them. In case the thickness κ^{-1} of the double layer around spherical colloids is small compared to their radius, overlapping double layers of two colloids can be considered as being essentially flat. An explicit expression for the interaction potential between two flat interfaces is calculated from the Helmholtz free energy in subsection 5.2.

5.1 The free energy of an electric double layer

In order to calculate the interaction potential between two charged colloids, we need an expression for the Helmholtz free energy instead of just the potential energy, since the ion-distribution in the double layers adjust to the distance between the two colloids. The Helmholtz free energy is defined as $\mathcal{F} = E - TS$, where E is the internal energy (potential energy due to inter-particle interactions plus kinetic energy), and S is the entropy. When an externally imposed (infinitesimally) small change is imposed on a system, like a small change of the distance between the two colloids, the change in the internal energy, according to the first law of thermodynamics, is equal to $dE = dw + dq$, where dw is the work that is involved, and dq is the heat-exchange between the reservoir with a fixed temperature in which the colloids are embedded. According to the second law of thermodynamics, $dq = T\,dS$, so that $d\mathcal{F} = dE - T\,dS - S\,dT = (dw + T\,dS) - T\,dS - S\,dT = dw - S\,dT$, provided that the small change is performed very slowly, such that the system remains in internal equilibrium. Hence, for the fixed temperature of the entire system, we have $dw = d\mathcal{F}$. Since the interaction force between the two colloids is related to the work (since work=interaction force $\times$ distance), a calculation of the Helmholtz free energy leads to an expression for the interaction potential between the colloids.

Consider a colloidal particle with carboxyl ($-COOH$) groups attached to its surface. On dissociation of the carboxyl group, and H^+-ion goes into solution, while the dissociated group $-COO^-$ renders the surface negatively charged. The free energy is calculated by a gedanken experiment, where the carboxyl groups are constrained to dissociate step wise in very small portions, sufficiently slow such that during the up-charging of the surface the double layer is always internally in equilibrium. The system as a whole (surface plus double layer) is during charging out-of-equilibrium, and reaches equilibrium only after the final charge is reached. For simplicity the solvent is assumed to contain, besides H^+-ions, just a single species of mono-valent negative ions (like, for example, chlorine Cl^-).

On release of a small portion of H^+-ions from the colloid's surface, the ion-concentrations in a given volume element $d\mathbf{r}$ located within the double layer changes. The accompanied change in free energy of that volume element is connected to the change of the number of ions within the volume element. The work dw involved to change the number of ions in the small volume element is now the work that is necessary to add dN particles. Since this work is proportional to dN, so that we have $d\mathcal{F} = dw = \mu\,dN$, where μ is referred to as *the chemical potential*. Consider two neighbouring volume elements, which are mutually in equilibrium, and dN particles are moved from volume 1 to volume 2, say. Since the change of the number of particles in volume 2 is $-dN$, the total change in free energy is , $d\mathcal{F} = (\mu_1 - \mu_2)\,dN$. Since the free energy attains a minimum value in equilibrium (so that $d\mathcal{F}/dN = 0$), the two chemical potentials should be equal: $\mu_1 = \mu_2$. The chemical potentials $\mu(\mathbf{r})$ of all volume elements located at position $\mathbf{r}$ within the double layer are therefore equal during the slow charging of the colloidal particle,

$$\mu(\mathbf{r}) \;=\; \text{constant} \quad , \quad \text{for } \mathbf{r} \text{ within the double layer.} \tag{23}$$

The colloid will be charged, keeping the surface charge density uniform. Let σ' denote the surface charge density during the charging process, which starts at 0 and increases up to σ, after which the entire system (double layer plus the colloidal surfaces) is in equilibrium. Let $\delta\mathcal{F}(\mathbf{r} \mid \sigma') = \mu_+(\mathbf{r} \mid \sigma')\delta N_+(\mathbf{r}) + \mu_-(\mathbf{r} \mid \sigma')\delta N_-(\mathbf{r})$ denote the change in free energy of a volume element located at $\mathbf{r}$ on changing the surface charge density, where $\delta N_\pm$ is the change

of the number of positive and negative ions in the volume element under consideration. Here, we use δ's to indicate small changes within volume elements, and later d's for the changes of quantities summed over all volume elements. The chemical potential consists of two parts: the work related to changes of inter-particle interactions and kinetic energy, and the electric work needed to bring ions to a position with a local electric potential $\Psi(\mathbf{r}\,|\,\sigma')$. The first part is the "chemical part" that is independent of the local potential, which will be denoted as $\mu^c_{\pm}$, while the second contribution is equal to $\pm e\,\Psi$. Hence,

$$\begin{aligned}\delta\mathcal{F}(\mathbf{r}\,|\,\sigma') &= \Big[\mu^c_+(\mathbf{r}\,|\,\sigma') + e\,\Psi(\mathbf{r}\,|\,\sigma')\Big]\,\delta N_+(\mathbf{r}) \\ &+ \Big[\mu^c_-(\mathbf{r}\,|\,\sigma') - e\,\Psi(\mathbf{r}\,|\,\sigma')\Big]\,\delta N_-(\mathbf{r})\;.\end{aligned}$$

Since the double layer is in internal equilibrium, the chemical potential is the same for all volume elements (see eq.(23)). We can therefore take the chemical potential equal to that of volume elements at the surface of the colloidal particle,

$$\mu^c_{\pm}(\mathbf{r}\,|\,\sigma') \pm e\,\Psi(\mathbf{r}\,|\,\sigma') \;=\; \mu^c_{s,\pm} \pm e\,\Psi_s(\sigma')\;,$$

where the subscript "s" refers to the value at the surface. The total change $d\mathcal{F}_{dl}(\sigma')$ of the free energy of the double layer is thus found by integration over all volume elements, noting that the total number of negative ions remains unchanged (the subscript "dl" stands for "double layer"),

$$d\mathcal{F}_{dl}(\sigma') \;=\; \Big[\mu^c_{s,+}(\sigma') + e\,\Psi_s(\sigma')\Big]\,dN_+\;, \tag{24}$$

where,

$$d\mathcal{F}_{dl}(\sigma') \;=\; \int d\mathbf{r}\;\delta F(\mathbf{r}\,|\,\sigma') \quad , \quad dN_+ \;=\; \int d\mathbf{r}\;\delta N_+(\mathbf{r})\;.$$

Note that dN_+ is the total number of H^+-ions added to the double layer, which is minus the number of $-COOH$ groups that are allowed to dissociate. The change of the free energy of the surface of the colloidal particle is similarly related to the surface value $\mu^c_{bound,+}$ of the chemical part of H^+-ions bounded to a $-COO^-$ group, and the electric potential,

$$d\mathcal{F}_{surf}(\sigma') \;=\; -\,\Big[\mu^c_{bound,+}(\sigma') + e\,\Psi_s(\sigma')\Big]\,dN_+\;, \tag{25}$$

where the subscript "$surf$" refers to the surface of the colloidal particles. The change of the total free energy of the system as a whole is thus equal to,

$$d\mathcal{F} \;=\; d\mathcal{F}_{dl}(\sigma') + d\mathcal{F}_{surf}(\sigma') \;=\; \Big[\mu^c_{s,+}(\sigma') - \mu^c_{bound,+}(\sigma')\Big]\,dN_+\;. \tag{26}$$

We now use the relation $\mu = \mu^0 + k_BT\ln\{\rho\}$, for the dependence of the chemical potential on the concentration ρ of a dissolved substance. This relation can be found in any standard textbook on thermodynamics. In the present case this relation reads,

$$\mu^c_{s,+}(\sigma') \;=\; \mu^0_+ + k_BT\,\ln\rho_{s,+}(\sigma')\;,$$

where $\rho_{s,+}$ is the concentration of H^+-ions in solution at the colloidal surface. Hence,

$$d\mathcal{F} \;=\; \Big[\mu^0_+ - \mu^0_{bound,+}(\sigma') + k_BT\,\ln\rho_{s,+}(\sigma')\Big]\,dN_+\;.$$

In the simplest approximation, which is used in the classic Verwey-Overbeek theory [1], the chemical potential $\mu^0_{bound,+}(\sigma')$ is taken independent of the surface concentration of H^+-ions, that is, $\mu^0_{bound,+}(\sigma')$ is assumed to be independent of σ'. For $\sigma' = \sigma$ the system as a whole is in equilibrium, where the free energy attains a minimum value, so that $d\mathcal{F}(\sigma)/dN_+ = 0$. It thus follows from the above result that,

$$\mu^0_+ - \mu^0_{bound,+} \;=\; -\,k_B T\,\ln \rho_{s,+}(\sigma)\;. \tag{27}$$

This can be regarded as an equation of state for the adsorption of ions from solution onto the surface. It thus follows that,

$$d\mathcal{F} \;=\; k_B T\,\ln\left\{\frac{\rho_{s,+}(\sigma')}{\rho_{s,+}(\sigma)}\right\}\,dN_+\;.$$

Using eq.(6) for the bulk concentration, and noting that $d\sigma' = -e\,dN_+/A$, with A the surface area of the colloidal particle, this finally leads to the following expression for the free energy of the double layer per unit area,

$$\frac{\mathcal{F}}{A} \;=\; \frac{1}{A}\int_0^\sigma dF \;=\; \int_0^\sigma d\sigma'\,\Psi_s(\sigma') - \sigma\,\Psi_s(\sigma)\;. \tag{28}$$

This is the fundamental result on which the calculation of the interaction potential in the next subsection is based.

5.2 The interaction potential

There is an important consequence of the equation of state (27) when the interaction between two colloids is considered. According to eq.(6) the equation of state can be written as,

$$\mu^0_+ - \mu^0_{bound,+} + k_B T\,\ln \rho^0_+ \;=\; e\,\Psi_s(H,\sigma)\;, \tag{29}$$

where ρ^0_+ is the concentration of H^+-ions outside the double layer. Since in that case the left hand-side of eq. (29) is just a constant, the surface potential must be a constant, independent of H. In other words,

There can only be equilibrium when the surface potentials of two identical colloids are independent of their separation .

This implies that the surface charge density changes with the distance between the colloids, in order to keep the surface potential fixed,

$$\sigma \;\equiv\; \sigma(H)\;,\; \textit{such that}\;,\; \Psi_s(H,\sigma(H)) = \textit{constant}\;,\; \textit{provided that}\; \mu^0_{bound,+}(\sigma') \equiv \textit{constant}\,.$$

The force $F^I = -\,(1/A)dF(H)/dH$ (where the superscript "I" stands for "interaction") between two flat interfaces per unit area is thus equal to,

$$F^I \;=\; -\,\frac{1}{A}\,\frac{d}{d\,H}\left[\int_0^{\sigma(H)} d\sigma'\,\Psi_s(H,\sigma') - \sigma(H)\,\Psi_s\right]_{\Psi_s = \textit{constant}}\;. \tag{30}$$

Note that for $\Psi_s(H,\sigma')$ in the integral, σ' is an integration variable, which is obviously independent of H. The general form of $\Psi_s(H,\sigma')$ is $\Psi_s\, f(H,\sigma')$, with Ψ_s the prescribed surface potential (which is equal to the surface potential of each of the two identical colloids at infinite separation). The differentiation with respect to H in the integral is thus understood as a partial differentiation of $f(H,\sigma')$.
The right hand-side of eq.(30) can be formally rewritten as [2],

$$F^I \;=\; -\frac{1}{A}\,\frac{d}{d\,H}\,\left[\int_0^{\sigma} d\sigma'\;\Psi_s(H,\sigma')\right]_{\sigma=\textit{constant}}\;. \tag{31}$$

Now the surface charge density is formally taken independent of the separation between the colloids. This is nothing more than a formal result, which does not imply that the surface charge density is independent of the separation. It is always the surface potential that is independent of H, not the surface charge density.
It is of course possible to use more realistic expressions for the chemical potential $\mu^0_{bound,+}(\sigma')$, based on an analysis of the dissociation-association equilibrium for the bounded carboxyl groups (see, for example, Refs.[2, 15]).
Let us now consider the solution of the linearized Poisson-Boltzmann equation for two flat plates. As before, let H denote the distance between the two flat interfaces, and let z denote the distance from the mid-plane, in between the two plates. That is, $z=0$ at the mid plane between the two surfaces. The solution of the linearized Poisson-Boltzmann equation (8) is now a sum of two exponentials,

$$\Psi(z) \;=\; A\,\exp\{-\kappa\, z\} + B\,\exp\{+\kappa\, z\} \quad , \quad -\tfrac{1}{2}\,H \le z \le +\tfrac{1}{2}\,H\;.$$

From symmetry, $\Psi(z)=\Psi(-z)$, and hence $A=B$. When $z=\pm\frac{1}{2}H$, the potential is equal to the surface potential $\Psi_s(H)$, which is a function of the distance H between the two plates. Hence (with "the hyperbolic-cosine function" $\cosh\{x\}=(\exp\{+x\}+\exp\{-x\})/2$),

$$\Psi(z) \;=\; \Psi_s(H,\sigma)\,\frac{\cosh\{\kappa\, z\}}{\cosh\left\{\frac{1}{2}\,\kappa\, H\right\}}\;.$$

Just as for the single plate as in section 4.2, the surface potential can be connected to the surface charge density σ, with the result,

$$\Psi_s(H,\sigma) \;=\; \frac{\sigma}{\epsilon_s\,\kappa}\,\frac{\cosh\left\{\frac{1}{2}\,\kappa\, H\right\}}{\sinh\left\{\frac{1}{2}\,\kappa\, H\right\}}\;. \tag{32}$$

Substitution into eq.(31) gives $F^I = -dV(H)/dH$, where the interaction potential $V(H)$ is equal to,

$$V(H) \;=\; \frac{1}{2\,A}\,\frac{\sigma^2}{\epsilon_s\,\kappa}\,\frac{\exp\left\{-\frac{1}{2}\,\kappa\, H\right\}}{\sinh\left\{\frac{1}{2}\,\kappa\, H\right\}} \;\approx\; \frac{1}{A}\,\frac{\sigma^2}{\epsilon_s\,\kappa}\,\exp\left\{-\kappa\, H\right\}\;. \tag{33}$$

[2] From eq.(30) we have,

$$\begin{aligned} F^I \;&=\; -\frac{1}{A}\left[\frac{d}{d\,\sigma}\int_0^{\sigma(H)} d\sigma'\,\Psi_s(H,\sigma')\,\times\,\frac{d\,\sigma(H)}{d\,H} + \int_0^{\sigma(H)} d\sigma'\,\frac{\partial\,\Psi_s(H,\sigma')}{\partial\,H} - \frac{d\,\sigma(H)}{d\,H}\,\Psi_s(H,\sigma)\right] \\ &=\; -\frac{1}{A}\int_0^{\sigma(H)} d\sigma'\,\frac{\partial\,\Psi_s(H,\sigma')}{\partial\,H} \;=\; -\frac{1}{A}\,\frac{d}{d\,H}\left[\int_0^{\sigma} d\sigma'\;\Psi_s(H,\sigma')\right]_{\sigma=\text{constant}}\;. \end{aligned}$$

Here, we chose a constant offset, such that $V(H \to \infty) = 0$, while the approximation in the last equation assumes that $\kappa H > 1$. According to the discussion above, the interaction force is obtained by differentiating this expression for $V(H)$ with respect to H, regarding the surface charge density as a constant, after which the H-dependent value of σ in terms of the prescribed surface potential must be substituted, as given in eq.(32).
The calculation of the interaction potential between two spheres with thin double layers can be done through the so-called *Derjaguin approximation*, which will be discussed in **chapter B4**.

6 Electrode polarization

In an experiment where an alternating external electric field is applied, charges are applied to two electrodes by means of a function generator. Just as for the charged colloids, electric double layers will build up at these charged electrode surfaces. The double layers partially screen the external charges, thus lowering the actual electric field strength within the bulk of the suspension, away from the electrodes. This phenomenon is referred to as *electrode polarization*. In studying transport, dielectric polarization, and field-induced phase transitions, the relevant field strength is that in bulk, as this is the field that is experienced by the colloids. It is therefore necessary to correct the applied field strength for electrode polarization.
The difference with what is discussed above is that we now have an alternating charge on the electrodes. Since the ions in solution have a finite diffusivity, there will be a phase lag between the applied field and the double layer structure, and hence the field strength in bulk. Note that the double layers are fully developed at zero frequency, so that the externally applied charges are fully compensated. The electric field strength in bulk is now zero. In this section we discuss how the field strength within the bulk of the suspension differs from the applied field strength as a function of the frequency of the applied field.
Consider for simplicity a $1-1$ salt solution confined between two flat electrodes which are separated by a distance L. A spatially homogeneous alternating electric field $E_{ext} = E_0 \cos\{\omega t\}$ is applied, where E_0 is the field amplitude and ω the frequency. For other electrode geometries than flat, parallel electrodes, electro-osmotic flow can be important, which complicates the analysis considerably. For the two-plate geometry osmotic-flow is absent.
Within the Debye-Hückel approximation, we have $\rho_\pm \nabla\Psi \approx c\,\nabla\Psi$, where c is the salt concentration. Combining eq.(4) for the local charge density ρ in solution and the equation of motion (2) for the ion-concentrations, it is found that,

$$\frac{\partial\,\rho(\mathbf{r},t)}{\partial\,t} \;=\; D\,\left[\,\beta\,e^2\,c\,\nabla^2\Psi + \nabla^2\rho\,\right]\;,$$

where, as before, D is the diffusion coefficient for the positive and negative ions, which is assumed not to be too different for the $\pm$-ions. With the Poisson equation (3) a closed equation of motion for the charge density is found,

$$\frac{\partial\,\rho}{\partial\,t} \;=\; D\,\left[\,\frac{d^2}{dz^2} - \kappa^2\,\right]\,\rho\;, \tag{34}$$

where κ is the inverse Debye length in eq.(9). Here it is used that the charge density and potential only vary in the z-direction, in the direction perpendicular to the two electrodes, where z varies from $-L/2$ to $+L/2$. The boundary conditions to the equation of motion (34) and the

Poisson equation are,

$$\begin{aligned} \frac{d}{d\,z}\,\rho + \epsilon_s\,\kappa^2\,\frac{d}{d\,z}\,\Psi &= 0\ , \quad \text{for} \quad z = \pm\tfrac{1}{2}\,L\ , \\ \Psi(z=\tfrac{1}{2}\,L) - \Psi(z=-\tfrac{1}{2}\,L) &= -\,E_0\,\cos\{\omega\,t\}\,L\ , \end{aligned} \tag{35}$$

where the first boundary condition ensures that there are no ion-fluxes through the electrodes, and the second boundary condition expresses that a voltage $-E_0\,\cos\{\omega\,t\}\,L$ is imposed across the two electrodes.
The system of equations (34-35), together with the Poisson equation (3), are most conveniently re-formulated in complex quantities. Let ρ' and ρ'' denote the in-phase and -out-phase components of the local charge density, that is, $\rho(\mathbf{r},t) = \rho'(\mathbf{r}\mid\omega)\,\cos\{\omega t\} + \rho''(\mathbf{r}\mid\omega)\,\sin\{\omega t\}$. Defining the complex charge density $\tilde{\rho} = \rho' - i\,\rho''$, it is easily seen that the real part of $\tilde{\rho}\,\exp\{i\omega t\}$ is equal to $\rho'(\mathbf{r}\mid\omega)\,\cos\{\omega t\} + \rho''(\mathbf{r}\mid\omega)\,\sin\{\omega t\}$. Similarly introducing the complex potential $\tilde{\Psi}$ the system (34-35) and the Poisson equation can be rewritten as,

$$\begin{aligned} \left[\frac{d^2}{d\,z^2} - \tilde{\kappa}^2\right]\,\tilde{\rho} &= 0\ , \\ \frac{d^2}{d\,z^2}\,\tilde{\Psi} &= -\,\frac{\tilde{\rho}}{\epsilon_s}\ , \\ \frac{d}{d\,z}\,\tilde{\rho} + \epsilon_s\,\kappa^2\,\frac{d}{d\,z}\,\tilde{\Psi} &= 0\ , \quad \text{for} \quad z = \pm\tfrac{1}{2}\,L\ , \\ \tilde{\Psi}(z=\tfrac{1}{2}\,L) - \tilde{\Psi}(z=-\tfrac{1}{2}\,L) &= -\,E_0\,L\ , \end{aligned} \tag{36}$$

where the complex-valued Debye screening length $\tilde{\kappa}^{-1}$ is given by,

$$\tilde{\kappa}^2 = \kappa^2 + i\,\frac{\omega}{D}\ .$$

Taking the square root it is found that,

$$\tilde{\kappa} = \kappa\,\left[\,f(\Lambda) + i\,g(\Lambda)\,\right]\ ,$$

where,

$$\begin{aligned} f(\Lambda) &= \frac{1}{\sqrt{2}}\,\left[1 + \left[1+\Lambda^2\right]^{1/2}\right]^{1/2}\ , \\ g(\Lambda) &= \frac{1}{\sqrt{2}}\,\left[-1 + \left[1+\Lambda^2\right]^{1/2}\right]^{1/2}\ , \end{aligned}$$

with the dimensionless frequency Λ equal to, [3]

$$\Lambda = \frac{\omega}{D\,\kappa^2}\ .$$

Since the charge density is an odd function of z, the solution to the first equation of motion is,

$$\tilde{\rho}(z) = C_1\,\sinh\{\tilde{\kappa}\,z\}\ . \tag{37}$$

[3] Note that, $f^2(\Lambda) - g^2(\Lambda) = 1$, and $f(\Lambda)\,g(\Lambda) = \frac{1}{2}\,\Lambda$, which are useful relations in the calculation of real- and imaginary parts.

From the Poisson equation it now follows that,

$$\tilde{\Psi}(z) \;=\; -\frac{1}{\epsilon_s\,\tilde{\kappa}^2}\,\tilde{\rho} + C_2\,z\;,$$

with C_2 another integration constant. The two boundary conditions lead to,

$$\begin{aligned} C_1\,\tilde{\kappa}\left[1-\frac{\kappa^2}{\tilde{\kappa}^2}\right]\cosh\left\{\tfrac{1}{2}\,\tilde{\kappa}\,L\right\} + \epsilon_s\,\kappa^2\,C_2 &= 0\;,\\ -\frac{2\,C_1}{\epsilon_s\,\tilde{\kappa}^2}\,\sinh\left\{\tfrac{1}{2}\,\tilde{\kappa}\,L\right\} + C_2\,L &= -\,E_0\,L\;, \end{aligned}$$

where the amplitude E_0 is taken along the minus z-direction. These two algebraic equations lead to,

$$\begin{aligned} C_1 &= \frac{\epsilon_s\,\tilde{\kappa}^2\,L}{\tilde{\kappa}\,L\left[\frac{\tilde{\kappa}^2}{\kappa^2}-1\right]\,\cosh\{\frac{1}{2}\tilde{\kappa}\,L\}+2\,\sinh\{\frac{1}{2}\tilde{\kappa}\,L\}}\,E_0\;,\\ C_2 &= -\frac{\tilde{\kappa}\,L\,\left[\frac{\tilde{\kappa}^2}{\kappa^2}-1\right]\,\cosh\{\frac{1}{2}\tilde{\kappa}\,L\}}{\tilde{\kappa}\,L\left[\frac{\tilde{\kappa}^2}{\kappa^2}-1\right]\,\cosh\{\frac{1}{2}\tilde{\kappa}\,L\}+2\,\sinh\{\frac{1}{2}\tilde{\kappa}\,L\}}\,E_0\;. \end{aligned}$$

Introducing the dimensionless frequency,

$$\Omega \;=\; \kappa\,L\,\Lambda \;=\; \frac{\omega\,L}{D\,\kappa}\;, \tag{38}$$

which is much larger than Λ for the common situation that $\kappa L \gg 1$, we have,

$$\tilde{\kappa}\,L\left[\frac{\tilde{\kappa}^2}{\kappa^2}-1\right] \;=\; i\,\frac{\tilde{\kappa}}{\kappa}\,\Omega \;\approx\; i\,\Omega\,,$$

provided that $\Lambda \ll 1$. Within the bulk of the solution, away from the electrodes, the charge density is zero, so that it follows from eqs.(38,38) that the electric field strength in the bulk of the solution is equal to,

$$E_{bulk} \;=\; -\,C_2 \;=\; \frac{i\,\Omega}{2+i\,\Omega}\,E_0 \;=\; \left[\frac{\Omega^2}{4+\Omega^2}+i\,\frac{2\,\Omega}{4+\Omega^2}\right]\,E_0\;.$$

For large Ω, $E_{bulk} = E_0$, so that no electrode polarization occurs. For $\Omega \to 0$ on the contrary, $E_{bulk} = 0$, so that the double layers at the electrodes completely screen the imposed field, as expected. Note that electrode polarization is of importance only for $\Omega < 20$, which justifies the leading order expansion with respect to $\Lambda = \Omega/\kappa\,L \ll 1$. According to the above mentioned meaning of the complex notation, we thus find that,

$$E_{bulk} \;=\; \left[\frac{\Omega^2}{4+\Omega^2}\,\cos\{\omega\,t\}+\frac{2\,\Omega}{4+\Omega^2}\,\sin\{\omega\,t\}\right]\,E_0\;. \tag{39}$$

As expected, there is an in-phase component ($\sim\; cos\{\omega t\}$) and an out-phase component ($\sim sin\{\omega t\}$).
A important result from the above analysis is that,

Electrode polarization becomes less important when the distance between the electrodes is increased (keeping the frequency fixed),

which follows from the expression (38) for the dimensionless frequency Ω and the fact that the result in eq.(39) only depends on that frequency. This is used in measurements of, for example, the electrophoretic mobility, where a small frequency is needed to be able to perform a measurement of the velocity of a colloidal particle for an essentially time-independent electric field. By choosing a large electrode separation, one can apply small frequencies without having to worry about electrode polarization. The physical origin of the electrode-separation dependence is as follows. Suppose that L is doubled. In order to have the same externally applied field strength, the charge on both electrodes must be doubled (since the potential is proportional to the charge, and the electric field is the potential difference between the electrodes divided by their distance). The same bulk field strength is then obtained when the charge within the double layers is also doubled. Since this takes twice as long, the frequency at which the same bulk field strength is found is twice as low.
The electric field within the bulk can generally be written as,

$$E_{bulk} \;=\; E_{0,bulk}\,\cos\{\omega\,t+\varphi\} \;=\; E_{0,bulk}\,\left[\,\cos\{\varphi\}\,\cos\{\omega\,t\}-\sin\{\varphi\}\,\sin\{\omega\,t\}\,\right]\;,$$

where $E_{0,bulk}$ is the amplitude of the electric field within the bulk, and φ is the phase lag as compared to the applied field, which both depend on frequency. Comparing to eq.(39) we thus find that the so-called attenuation factor γ is equal to,

$$\gamma \;\equiv\; \frac{E_{0,bulk}}{E_0} \;=\; \sqrt{\left(\frac{\Omega^2}{4+\Omega^2}\right)^2+\left(\frac{2\,\Omega}{4+\Omega^2}\right)^2} \;=\; \frac{\Omega}{\sqrt{4+\Omega^2}}\;. \tag{40}$$

One way to experimentally verify this expression for the field-strength attenuation is to determine field-induced phase transitions as a function of L [16]. This can, for example, be done for suspensions of long and thin, highly charged colloidal rods. Fd-virus particles can be used as a model system for charged colloidal rods. These viruses consist of a DNA strand that is covered with 2700 coat proteins. Their length is $880\,nm$, the core-width is $6.8\,nm$, and the persistence length is about $2500\,nm$. The surface of the fd-virus particle is highly charged: it carries 10 (negative) elementary charges per nm. About $80-90\,\%$ of these bare charges are compensated by condensed ions. Starting with a suspension that is in isotropic-nematic coexistence, several phases can be induced by a low frequency alternating electric field. The phase diagram is given in Fig.3a. The phase that is referred to as the N-phase is an isotropic-nematic coexistence, the N^*-phase is a chiral nematic, the D-state is a dynamical state where small nematic domains persistently melt and form, while the H-phase is a uniform homeotropic phase, where the rods are aligned along the field direction. The images in Fig.3a are images of the morphology of the various phases/states taken between crossed polarizers. As the fd-virus particles are birefringent, the white regions correspond to local orientational ordering. The frequency at which charge polarization is essentially absent can be estimated to be about $500\,Hz$. The H-phase is therefore believed to be stabilized by hydrodynamic interactions through electro-osmotic flow induced within the double layers and the layer of condensed ions [16]. The dynamical state is due to the cyclic dissociation and association of condensed ions [17]. For a given frequency, field-induced transitions from one state to the other obviously occur at a fixed field strength within the bulk, as this is the field strength that is experienced by the colloidal rods. The applied field strength where a transition is observed may differ, however, as the distance between the electrodes is varied. The L-dependence of the applied field strength where the transitions from the N^* to the dynamical state D, and from the N to the N^* phase are observed is plotted in Fig.3b. The solid lines are fits of the data to eq.(40) with two fitting parameters: the

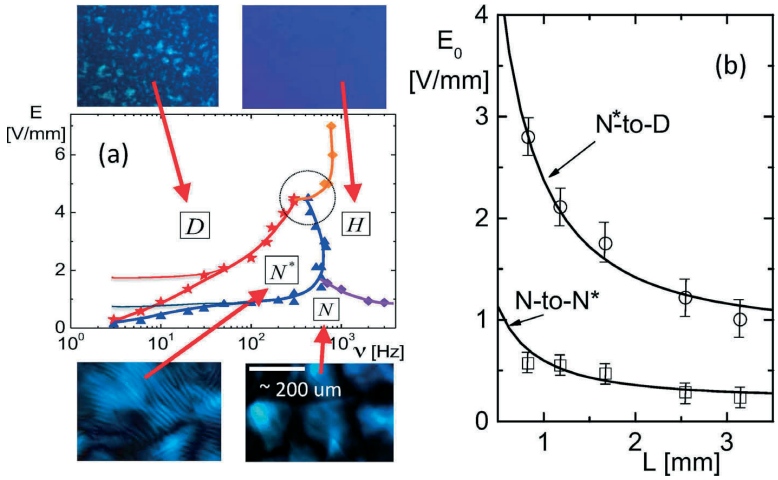


Fig. 3: *(a) The phase/state diagram in the electric field strength versus frequency plane, for a suspension of fd-virus particles that is in isotropic-nematic coexistence without the electric field [16]. The various phases and states are briefly discussed in the main text. The images are taken with the sample in between two crossed polarizers, where bright regions correspond to local orientational order. The solid lines at low frequencies are the transition lines as probed without correction for electrode polarization. (b) The applied field strength where the two transitions N-to-N^* and N^*-to-D are observed as a function of the distance between the electrodes. The solid lines are fits to eq.(40) [16].*

constant value of $E_{0,bulk}$ where the transitions occur (which is of course different for the two transitions), and the diffusion coefficient D of the ions (which is the same for the two transitions). As can be seen the fits describe the data quite well, while the diffusion coefficient that is found is $D = 2.1 \times 10^{-9}\, m^2/s$, which is well within the range of $1 - 3 \times 10^{-9}\, m^2/s$ known for ion-diffusion coefficients. The two solid red and blue lines in Fig.3a are the measured transition lines without correction for electrode polarization. The true bulk field strength is given by the corresponding two lower curves.

7 Electrophoretic Mobility

The surface potential of colloidal particles can be determined experimentally through the velocity that the particles attain under the action of an external electric field. This velocity is referred to as *the electrophoretic velocity*. For sufficiently small electric field strengths E, the electrophoretic velocity v is proportional to the field strength. The proportionality constant,

$$\mu_m \;=\; \frac{v}{E}\,,$$

is referred to as *the electrophoretic mobility*. In this section we consider the relation between the surface potential of the colloidal particles and their electrophoretic mobility for two limiting cases, where the double thickness is either very large or small compared to the radius of the spherical colloid. Within the Gouy-Chapman approach which is adopted throughout this lecture, the difference between the potential at the shear-plane (as discussed in the introduction) and the

static surface potential is neglected. Furthermore, we assume low surface potentials, so that the Debye-Hückel approximation can be used to simplify the gouverning equations. In addition we assume that the core of the spherical colloid is a non-conducting dielectric, with a dielectric constant that is similar to that of the solvent, so that the external electric field is not perturbed by the mere presence of the colloidal core.

There is a large body of literature on electrophoresis, where theories for arbitrary Debye lengths have been discussed, where the effect of surface conductivity is included, where large surface potentials are considered, and where finite ion-size effects and hydrodynamic slip phenomena are accounted for. Some of the seminal papers in this area are Refs.[18, 19, 20, 21, 22, 23] (which is by no means a complete list of references), while the book of Hunter [2] contains an in-depth analysis of electrophoresis. The larger part of this work relies on numerical solutions of the relevant differential equations for the charge distribution and fluid flow.

A large distance between electrodes is used in electrophoretic experiments, so that a constant electric field can be applied during a sufficiently long time needed to measure the colloid velocity. After that time the direction of the applied field is reversed, in order to prevent electrode polarization (see the previous section).

Once the colloidal particle attained a constant velocity (after typically a nanosecond), the total (thermally averaged) force on the core of the colloid is zero. As sketched in Fig.4b, there are four forces to be distinguished: (i) the friction force $-\zeta\, v$ with the fluid, where ζ is the friction coefficient, (ii) the force $Q\, E$ of the eternal field on the total bounded charge Q on the surface of the colloid, (iii) the so-called *retardation force* F_{def}, that is, the force of the deformed electric double layer on the colloid's surface charges, and (iv) the electro-osmotic force F_{osm}. The deformation of the double layer has two origins: it is due to the force that the external electric field exerts on volume elements within the double layer, as well as the friction forces on ions due to the fluid flow around the core of the colloid as it moves through the solvent. The force that the electric field exerts on ions within the double layer is transferred to the solvent, which is thus set into motion. This flow is referred to as *electro-osmotic flow*, which exerts an additional hydrodynamic friction force F_{osm} on the core of the colloid, and is referred to as *the electro-osmotic friction force*. Both forces F_{def} and F_{osm} are some times referred to as *retardation forces*, as they both result from the deformation of the double layer. Hence,

$$0 \;=\; -\zeta\, v + Q\, E + F_{def} + F_{osm} \quad \rightarrow \quad \mu_m \;=\; \frac{1}{\zeta\, E}\left[\, Q\, E + F_{def} + F_{osm}\,\right] . \tag{41}$$

Here we consider two limiting cases, where the double layer thickness is either very large or small compared to the radius a of the spherical colloidal particle.

For a very large Debye length, such that $\kappa\, a \ll 1$, the ion-concentration within the double layer tends to zero, so that the colloid can be regarded as a charged sphere embedded in an uncharged dielectric medium, that is, the pure solvent. In that case both the electro-osmotic force as well as the retardation force are both zero. It thus follows immediately from eq.(41) that (with η the shear viscosity of the solvent),

$$\mu_m \;=\; \frac{Q}{\zeta} \;=\; \frac{Q = 4\,\pi\,\epsilon_s\, a\, \Psi_s}{\zeta = 6\,\pi\,\eta\, a} \;=\; \frac{2}{3}\,\frac{\epsilon_s\, \Psi_s}{\eta} \quad , \quad \kappa\, a \ll 1\;, \tag{42}$$

where eq.(11) that connects the charge Q with the surface potential Ψ_s has been used. This is the Hückel formula for the electrophoretic mobility for thick double layers.

For a very thin double layer, such that $\kappa\, a \gg 1$, the variation of the solvent velocity and the potential in the radial direction is much more pronounced as compared to their variations along

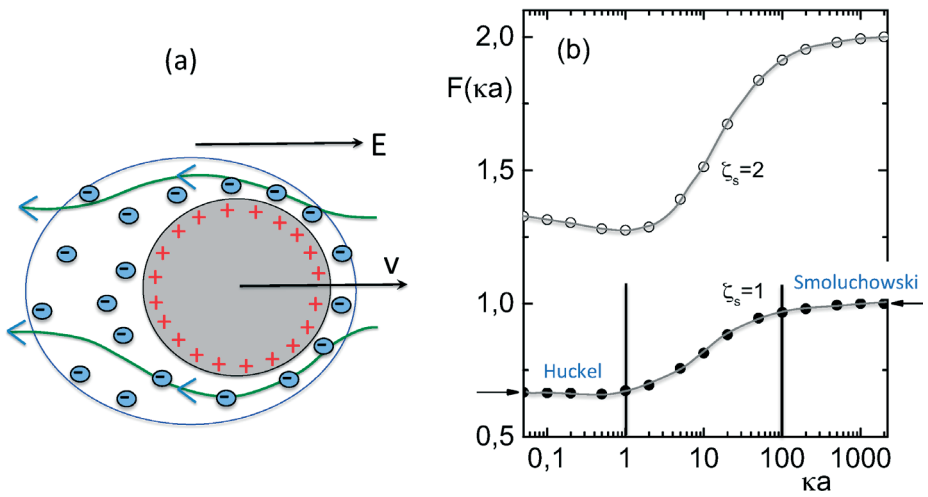


Fig. 4: *(a) A sketch of the double-layer deformation (indicated by the blue line) and the flow around the sphere in an external electric field. (b) The dimensionless electrophoretic mobility* $F(\kappa a) = \mu_m \eta/\epsilon_s \Psi_s$ *as a function of* $\kappa\, a$*, based on numerical results from Ref.[18], for two values of the dimensionless surface potential* $\zeta_s = e\,\Psi_s/k_B\,T$*, as indicated in the figure. The arrows indicate the limiting Hückel and Smoluchowski values for the electrophoretic mobility.*

the surface. The surface can thus be regarded as locally flat, where only spatial derivatives perpendicular to the surface are of importance. The solvent velocity $\mathbf{u}$ obeys the so-called *Navier-Stokes equation* for low Reynolds numbers (see **chapter B3**),

$$\mathbf{0} \;=\; \eta\nabla^2\mathbf{u} - \nabla p - \rho\,\nabla\Psi \;. \tag{43}$$

This equation expresses force-balance, where $\eta\nabla^2\mathbf{u}$ accounts for friction forces between sliding layers of fluid, $-\nabla p$ is the force due to gradients in the pressure p, and the last term is due to electric forces. The electrophoretic mobility can be found without explicitly solving the Navier-Stokes equation, by considering the flow near to the surface of the colloidal sphere where the electric field is parallel to the surface. Within a reference frame that is moving along with the colloid and for a flat surface we have, $\nabla^2\mathbf{u} = d^2\mathbf{u}/dr^2$, where r is the spatial variable in the radial direction, perpendicular to the surface. For a flat surface the pressure is a constant, so that $\nabla p = \mathbf{0}$. The ion-concentration and electric potential are written as,

$$\begin{aligned} \rho(\mathbf{r}) &= \rho^{(0)}(\mathbf{r}) + \rho^{(1)}(r) \;, \\ \Psi(\mathbf{r}) &= \Psi^{(0)}(\mathbf{r}) + \Psi^{(1)}(r) - \mathbf{E}\cdot\mathbf{r} \;, \end{aligned}$$

where $\rho^{(0)}$ and $\Psi^{(0)}$ are the unperturbed ion-charge density and potential, respectively, that is, the charge density and potential without the external field, while $\rho^{(1)}$ and $\Psi^{(1)}$ are perturbations due to the external field. Furthermore $\mathbf{E}$ is the external field. Within the Debye-Hückel approximation, products like $\rho^{(0)}\,\Psi^{(0)}$ and $\rho^{(0)}\,\Psi^{(1)}$ can be neglected, being of second order in the charge of the colloid. To linear order in the external field strength, the only contribution of interest in the last term in eq.(43) is therefore $\rho^{(0)}\,\mathbf{E}$. The Navier-Stokes equation thus reduces to,

$$\eta\,\frac{d^2\mathbf{u}(r)}{dr^2} \;=\; -\,\rho^{(0)}(r)\,\mathbf{E} \;=\; \epsilon_s\,\mathbf{E}\,\frac{d^2\Psi^{(0)}(r)}{dr^2} \;,$$

where in the last equation the Poisson equation has been substituted. Two integrations thus leads to,

$$\eta\,\mathbf{u}(r) \;=\; \mathbf{C}_0 + \mathbf{C}_1\,r + \epsilon_s\,\mathbf{E}\,\Psi^{(0)}(r)\;. \tag{44}$$

The integration constants can be determined from the boudary conditions (remember that the frame of reference is moving along with the colloidal particle with the electrophoretic velocity $\mathbf{v}$),

$$\begin{aligned} \mathbf{u} &= -\mathbf{v} \quad , \quad \text{for } r\to\infty\;, \\ \mathbf{u} &= \mathbf{0} \quad , \quad \text{for } r = a\;. \end{aligned}$$

The first boundary condition implies that $\mathbf{C}_1 = \mathbf{0}$ and $\mathbf{C}_0 = -\eta\,\mathbf{v}$. The second condition implies that $\mathbf{C}_0 = -\epsilon_s\,\mathbf{E}\,\Psi_s^{(0)}$. It thus follows that $\eta\,\mathbf{v} = \epsilon_s\,\mathbf{E}\,\Psi_s$, where Ψ_s is understood to be the unperturbed surface potential. Hence,

$$\mu_m \;=\; \frac{\epsilon_s\,\Psi_s}{\eta} \quad , \quad \kappa\,a \gg 1\;. \tag{45}$$

This is the Smoluchowski formula for the electrophoretic mobility.
The electrophoretic mobility differs a factor of $2/3$, depending whether there is a thick or thin double layer. For finite double-layer thicknesses, the electrophoretic mobility is a monotonically increasing function of $\kappa\,a$. The calculation of the electrophoretic mobility for such intermediate double-layer thicknesses is quite complex (see, for example Refs.[2, 18]). The dimensionless electrophoretic mobility $F(\kappa a) = \mu_m\,\eta/\epsilon_s\,\Psi_s$ is plotted in Fig.4b as a function of $\kappa\,a$. The data points in this plot are numerical results taken from Ref.[18] (their table I on page 84), for two values of the dimensionless surface potential $\zeta_s = e\,\Psi_s/k_B\,T$. For $\zeta_s = 1$ the Debye-Hückel approximation considered above is reasonably accurate. The arrows on the vertical axes in Fig.4b indicate the Hückel and Smoluchowski limits discussed above, which agree very well with the numerical results. For larger surface potentials there are large deviations from the Debye-Hückel approximation, as can be seen from the plot for $\zeta_s = 2$ in Fig.4b. The double-layer thickness dependence of the electrophoretic mobility becomes non-monotonic for large surface potentials with a minimum value around $\kappa a \approx 1$. The minimum becomes much more pronounced for even larger values of the surface potential.

References

[1] E.J.W. Verwey, J.Th.G. Overbeek, *Theory of the Stability of Lyotropic Colloids*, Dover Publications (1999)

[2] R.J. Hunter, *Zeta Potential in Colloid Science: Principles and Applications*, Academic Press (1988)

[3] J. Israelachvili, *Intermolecular & Surface Forces*, Academic Press (1991)

[4] J. Lyklema, *Fundamentals of Interface and Colloid Science. Volume I: Fundamentals*, Academic Press (1993)

[5] E. Waisman, J.L. Lebowitz, J. Chem. Phys. **56**, 3093 (1972)

[6] N. Grønbech-Jensen, K.M. Beardmore, P. Pincus, Physica A **261**, 74 (1998)

[7] E. Trizac, J.-L. Raimbault, Phys. Rev. E **60**, 6530 (1999)

[8] R. Kjellander, A.P. Lyubartsev, S. Marcelja, J. Chem. Phys. **114**, 9565 (2001)

[9] R. Okamoto, A. Onuki, J. Chem. Phys. **131**, 094905 (2009)

[10] C. Contreras-Aburto, G. Nägele, J. Chem. Phys. **139**, 134109 (2013)

[11] C. Contreras-Aburto, G. Nägele, J. Chem. Phys. **139**, 134110 (2013)

[12] G.S. Manning, J. Chem. Phys., **51**, 924 (1969)

[13] G.S. Manning, J. Chem. Phys., **51**, 3249 (1969)

[14] B.H. Zimm, M. Le Bret, J. Biom. Struc. Dyn., **1**, 461 (1983)

[15] D.Y.C. Chan, D.J. Mitchell, J. Coll. Int. Sci. **95**, 193 (1983)

[16] K. Kang, J.K.G. Dhont, Soft Matter **6**, 273 (2010)

[17] K. Kang, J.K.G. Dhont, Soft Matter **10**, 1987 (2014)

[18] P.H. Wiersema, A.L. Loeb, J. Th.G. Overbeek, J. Coll. Int. Sci. **22**, 78 (1966)

[19] R.W. O'Brien, L.R. White, J. Chem. Soc.Far. Trans 2, **74**, 1607 (1978)

[20] R.W. O'Brien, R.J. Hunter, Can. J. Chem., **59**, 1878 (1981)

[21] M. Fixman, J. Chem. Phys. **78**, 1483 (1983)

[22] J. Ennis, J.L. Anderson, J. Coll. Int. Sci. **185**, 497 (1996)

[23] A.S. Khair, T.M. Squires, Physics of Fluids **21**, 042001 (2009)

F 4 Introduction to thermal gradient related effects

S. Wiegand
Institute of Complex Systems / Soft Condensed Matter
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Temperature gradients: experimental observations **4**
2.1 Effects in solid materials 4
2.2 Effects in gases 5
2.3 Effects in liquid mixtures and Soft matter 5

3 Interesting materials designed for applications in temperature gradients **9**
3.1 Gold particles in polymer networks 9
3.2 Janus Particles 10
3.3 Is it possible to build thermophoretic machines? 10

4 Commercial Instruments utilizing temperature gradients **12**
4.1 Field-flow fractionation 12
4.2 MicroScale Thermophoresis 15

5 Outlook **15**

A Theoretical description of thermal diffusion **16**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

In a three-dimensional Cartesian coordinate system the gradient, ∇f, of a scalar field, $f(x, y, z)$ is given by

$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k} \tag{1}$$

where **i**, **j**, **k** are the standard unit vectors. ∇f points in the direction of the greatest rate of increase of the scalar field, f, and its magnitude is the slope of the graph in the direction of ∇f. In the case of a scalar temperature field, T, we can measure at each point (x,y,z) a temperature $T(x, y, z)$ assuming that the temperature does not change with time. The temperature gradient, ∇T points in the direction, where the temperature shows the largest increase. In SI units ∇T is expressed in Kelvin per meter (K/m).
As summarized in figure 1 temperature gradients can occur in gases, fluid systems and solid materials. While the scenario for liquids systems is still quite ambiguous due to the lack of a microscopic theory the temperature gradient effects in gases and solids are much better understood and can be described by a kinetic and atomic theory, respectively. Often experiments with temperature gradients are accompanied by convection, which occurs for certain spatial configurations as illustrated in figure 1. If the system is heated from above or if very narrow channels with a height of a few microns are used, the inset of convection can usually be avoided. The focus of this lecture will be on liquid and complex soft matter systems.

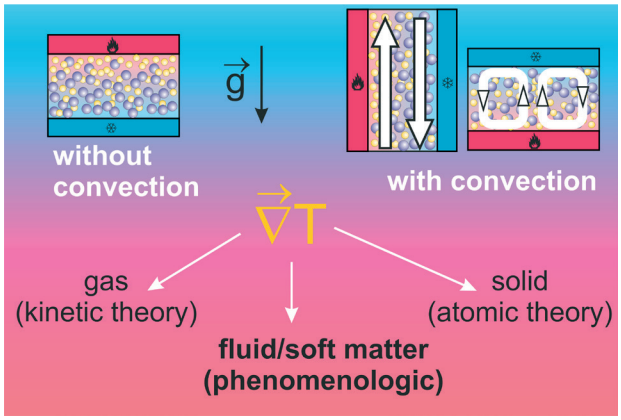


Fig. 1: *Temperature gradients occur in gases, fluids and solid materials. Depending on the spatial distribution of the temperature gradient in a gravity field $\vec{g}$ free convection occurs.*

In general a temperature gradient can cause a mass flux in a multicomponent system. This effect, first observed by Ludwig [1] and later systematically investigated by Soret [2], is called thermodiffusion, Ludwig-Soret effect or thermophoresis. For liquids and soft matter there is only a phenomenological description by the Onsager reciprocal equation (c.f. Section A), while the effect in gases and solids can be described by the kinetic gas [3] and atomic theory [4], respectively. Although there is no microscopic theory available most of the thermophoretic applications are for liquids or more complex soft matter systems.
Practically temperature gradients occur in many different areas such as meteorology and geology, but also process engineering, analytical techniques and living systems. For instance

the spatial distribution of the different crude oil components in petroleum reservoirs is influenced by thermal gradients [5]. Stationary models to mimic the compositional distribution of the reservoir are extended and consider thermal effects as a small steady flux to find a better agreement between the model and reality. It turns out that human sperms are very sensitive to temperature gradient as low as 0.014 K/mm directing the sperm to the warmer fertilization site [6]. Note that the effect here is *thermotaxis* and not *thermophoresis*. Recently Okabe et al. [7] reported a 2 K higher temperature in the nucleus compared to the cell plasma. Note that this result has been questioned recently [8]. There are analytical methods such as the Thermal Field flow fractionation (th-FFF) technique [9] for separation and characterization of synthetic soft matter systems. Recently the Microscale Thermophoresis (MST) [10] has been developed to monitor biochemical reactions utilizing thermophoresis. Also in the context of molecular evolution thermal gradients are discussed [11] and it was possible to duplicate DNA strands with a device combining thermal diffusion and convection [12].

Due to the fact that temperature gradients occur in many different research fields the nomenclature of the various (diffusion) coefficients is often quite confusing, so that we briefly give a summary of the used symbols and their meaning. We denote the ***thermal diffusivity***, D_{th}, which is the diffusion coefficient governing the dynamics of heat dissipation in the three-dimensional heat equation

$$\frac{\partial T\left(x,y,z,t\right)}{\partial t} = D_{\mathrm{th}}\nabla^2 T\left(x,y,z,t\right). \tag{2}$$

It can be expressed by the thermal conductivity κ, the density, ρ, and the specific heat, c_p, whereas $D_{\mathrm{th}} = \kappa/\rho c_p$ holds. D is the normal (ordinary, mass, translational, collective) ***diffusion coefficient*** associated with the diffusion of the solute (molecule, polymer, colloid). Assuming a constant D the diffusion is governed by the diffusion equation

$$\frac{\partial c\left(x,y,z,t\right)}{\partial t} = D\nabla^2 c\left(x,y,z,t\right), \tag{3}$$

where $c(x,y,z,t)$ is the concentration of the diffusing component. Like D_{th}, it is usually measured in (m^2/s). The ***thermal diffusion coefficient*** D_{T}, has the units (m^2/(s K)) and is defined by an extension of the isothermal diffusion equation

$$\frac{\partial c\left(x,y,z,t\right)}{\partial t} = D\nabla^2 c\left(x,y,z,t\right) + D_{\mathrm{T}} c\left(x,y,z,t\right)\left[1-c\left(x,y,z,t\right)\right]\nabla^2 T\left(x,y,z,t\right). \tag{4}$$

D_{T} is a measure for the strength of the coupling among temperature gradient, as the driving force, and mass flux due to the Ludwig-Soret effect (further details are given Section A). Finally, the quantity $S_{\mathrm{T}} = D_{\mathrm{T}}/D$ is termed Soret coefficient (Ludwig-Soret coefficient, or thermal diffusion ratio) and is measured in (K^{-1}). S_{T} is proportional to the ratio $\Delta c/\Delta T$ with the established concentration change, Δc and the temperature difference, ΔT. Note, that this holds only, if the temperature varies linearly and the temperature gradient is constant. The sign of the S_{T} (and also D_{T}) is positive, if the first named component enriches at the cold side and negative otherwise. Usually, one would specify S_{T} for the component that has been used for the definition of the concentration. Note that this definition does not depend on the densities of the two components.

The chapter is organized as follows. In the following we will give a brief overview, what is known about thermal gradients in solid materials, gaseous, liquids and soft matter systems. The focus will be on experimental observations made for liquids and soft matter systems. Due to the lack of a microscopic understanding of the effect in liquids, we will only be able to give

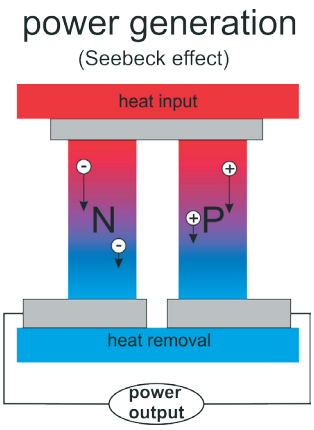


Fig. 2: *Illustration of the Seebeck effect. When heat flows across the junction, electrical current is generated.*

some rules of thumb. In the following section we will present some materials, which show or might show interesting effects in temperature gradients. And in the final section we discuss two commercial instruments utilizing temperature gradients. In the appendix we will give a heuristic description of the thermal diffusion phenomena.

2 Temperature gradients: experimental observations

2.1 Effects in solid materials

In solid materials two effects in temperature gradients are observed. One is the so-called ***Seebeck effect*** (the inverse of the *Peltier effect*), which is associated with the generation of a voltage along a conductor when it is subjected to a temperature gradient [13]. Another effect is the *thermomigration*, which is identical with Soret effect in alloys and semiconductors and denotes the movement of components in a temperature gradient.

The **Seebeck effect** occurs, when dissimilar metals, which are electrically in series and thermally in parallel connected, are exposed to a temperature gradient. Under these conditions a electric current is observed. As material often p - and n-doped semiconductors are used. Note that the electric current is propagated by electrons in n-type materials and by holes (traveling in the opposite direction) in p-type materials. As shown in Fig. 2.1, if the junctions at the top are heated and those at the bottom are cooled (producing a temperature gradient), electron/hole pairs will be created across the p-n-junction at the hot end and will absorb heat in this process. Electrons will flow away from the junction in the n-type material, and holes will flow away in the p-type material. The pairs recombine and reject heat at the cold ends. A voltage potential, the Seebeck voltage, which drives the hole/electron flow, is created by the temperature difference between the hot and cold ends of the thermoelectric elements. The net voltage appears across the bottom of the thermoelectric element legs. The Seebeck effect forms the basis of the operation of thermoelectric couples (thermocouples) used extensively in temperature-measurement systems. Electrical connections can be made from the thermoelectric device to an external load to extract power.

Using this effect scientists try to harvest electricity from waste heat, which should improve the cooling of highly concentrated microelectronic devices. **Thermomigration** can be used for doping semiconductors [14], but often the effect causes degradation problems during the operation of semiconductor devices [15] due to very high temperature gradients in the order of 100 K/mm.

2.2 Effects in gases

The first technical use of the thermodiffusion was the isotope separation of gas mixtures [16]. Clusius and Dickel [17] used in 1939 a 36 m high thermogravitational column to separate chlorine in its isotopes by combining thermal diffusion and convection. At that time this device was the most effective device for this purpose and the Americans used similar columns, in the so-called Manhattan Project, to enrich uranium hexafluoride. In the first half of the 20th century especially isotope mixtures consisting of hydrogen, deuterium and tritium, nobel gas mixtures and small polyatomic molecules such as carbon disulfide and ammonia have been investigated as function of concentration, temperature and pressure. Most of these early experiments and their theoretical analysis have been summarized in a book chapter by Waldmann [18]. For the theoretical description of the thermophoretic results of gaseous mixtures kinetic theories are used, which has been developed for monoatomic gases [19, 20]. The key elements of these early studies are the following: (i) If the mass difference between the two components is large, the heavier component enriches at the cold side [21]. (ii) For binary gas mixture with a small mass difference sign changes of the Soret coefficient with temperature and concentration are observed. Especially the second point has fallen into oblivion.

Nowadays there are no noteworthy practically applications of thermodiffusion in gaseous. From the theoretical point of view gaseous systems are interesting because they can be described a microscopic theory, and the calculated values have a higher accuracy than the experimentally determined values [22].

2.3 Effects in liquid mixtures and Soft matter

Especially in the last 20 years numerous systematic studies of binary liquid mixtures, polymer solutions and colloidal suspensions have been performed and in the last five years also biopolymers and colloids are in the focus of interest. A phenomenological description of thermal diffusion can be found in the appendix A. So far there is no microscopic picture to predict the direction of the thermophoretic motion and even an attempt to summarize the individual results into some roadmap as done in figure 2.3 looks quite complicated. In the following we will slowly go through the roadmap and summarize a few experimental works and theoretical concepts supporting the rules of thumb.

The first three properties ***mass***, ***moment of inertia*** and ***size*** are also essential part of the kinetic theory for gases. Typically the heavier or denser component, the component with the larger moment of inertia and the larger diameter moves to the cold side [23]. Note that principle *the-heavier-mass-goes-to-the-cold* can be overwritten, if some of the other effects, which lead to a thermophilic behavior is stronger this occurs often, if the mass difference is not so large. But even in the case of polymer in a solvent the statement: *"The movement of macromolecules in a temperature gradient is always in the direction from the hot to the cold region"* [24] does not always hold [25].

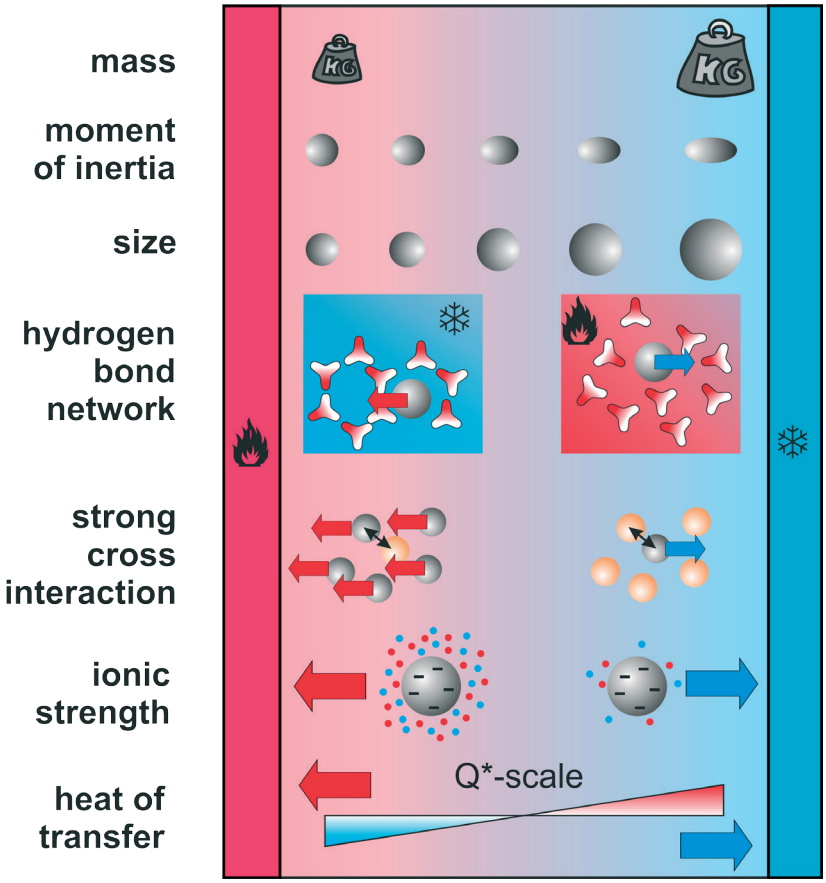


Fig. 3: *Sketch of a roadmap to identify the direction of the thermophoretic motion. Only general trends are identified, which do not hold for all but the majority of liquid systems. For instance* ***mass****,* ***moment of inertia*** *and* ***size*** *often govern the behavior of gaseous systems and are considered in the kinetic theory of gases.* ***Hydrogen bonds*** *and* ***cross interactions*** *are especially important for systems containing water. Charge effects or the* ***ionic strength*** *dependence are discussed for salt solutions and colloidal mixtures. Finally, the importance of the* ***heat of transfer****-scale, Q*-scale, will be demonstrated for low molecular mixtures. Further details are in the text.*

If the molecules are non-polar some concepts could be confirmed systematic isotope substitutions. From the measurements of cyclohexane and its isotope substitutes mixed with other organic compounds, two distinct contributions to the Soret coefficient could be identified. In addition to the contribution stemming from the difference between the molar mass ΔM and the moment of inertia ΔI [26, 27], the chemical differences between the two components also used to be considered [28, 29]:

$$S_T = S_T^0 + a_M \Delta M + b_I \Delta I \tag{5}$$

where S_{T}^0 is shown to be concentration dependent, while the part from the isotopic effect does not depend on concentration. The study of $n-$alkane in organic solvent mixtures shows the opposite moment of inertia trend [30], because with increasing degree of branching the heptane isomers have a stronger tendency to go to the cold side.

For aqueous mixtures and solutions the influence of ***hydrogen bond*** formation and specific interactions dominate the behavior. This leads to a universal temperature dependence of the Soret coefficient, which has been found for many aqueous systems [31, 32]. Often the temperature dependence of the the Soret coefficient can be described by an empirical law [31],

$$S_{\mathrm{T}}(T) = S_{\mathrm{T}}^{\infty}\left[1 - \exp\left(\frac{T^* - T}{T_0}\right)\right] \tag{6}$$

with the parameters S_{T}^{∞}, T^* and T_0. Recently, Wang *et al.* [33] explained the physical mechanism behind this universal temperature dependence and related it to the formation of hydrogen bonds in aqueous solutions. They adopt a free energy minimization concept, which is often used to explain the closed loop phase diagrams under isothermal conditions to this non-isothermal condition. Strictly speaking, a free energy is not defined under non-isothermal conditions. However, experimental and theoretical studies have shown that local thermodynamic equilibrium can be successfully applied to describe thermophoresis [34, 35]. They argue that the system locally minimizes its free energy $F = U - TS$ by the formation of hydrogen bonds, so that at lower overall temperature the water molecules accumulate on the cold side and the solute molecules on the warm side. At high overall temperatures the energy gain due to the formation of hydrogen bonds becomes less important, therefore the system minimizes its free energy by maximizing the orientational entropy. Thus, it is favorable for the small water molecules to be on the warm side in order to maximize their orientational and translational entropy. This hypothesis is also supported by the observation, that adding of a component, which is opening of hydrogen bonds, shows the same effect as an increase in temperature [36].

Another observation is, that in the presence of strong ***cross interactions*** always the minority component accumulates on the cold side, which leads to a sign change with concentration. The physical picture behind this observation is that if the cross interactions, ε_{12} are larger than the interactions between the pure compounds, ε_{11} and ε_{22} the minor component, for example 1, will tend to concentrate in the cold region because this will lower the total energy of the system. When component 1 is very dilute, only the cross interaction ε_{12} will compete with the pure interaction (ε_{22}) of the other constituent 2. On the other hand, if the concentration of 1 increases to the point to be the majority, then the constituent 2 will also tend to concentrate in the cold region for the same energetic argument. In this case, the thermal diffusion coefficient changes its sign when the concentration of the mixture is changed between these two limiting cases [23].

The ***ionic strength*** dependence of charged spherical and rod-like colloids has been successfully described by several papers of the Dhont group [37, 38, 39]. The Soret coefficient of a single

charged colloid is given by,

$$S_{\mathrm{T}} = \frac{D_{\mathrm{T}}}{D} = \frac{1}{T}\{1 + \frac{1}{4}(\frac{4\pi l_B^2 \sigma}{e})^2 \frac{1}{(1+\kappa R)^2} \frac{\kappa R^4}{l_B^3}[1 - \frac{dln\epsilon}{dlnT}(1 + \frac{2}{\kappa R})]\} + A(T)\ , \qquad (7)$$

where $\sigma = Q/4\pi R^2$ is the surface charge density, $l_{\mathrm{B}} = \beta e^2/4\pi\epsilon$ is the Bjerrum length (0.71nm for water at room temperature), with ϵ the dielectric constant of water at the ambient temperature, where κ is the reciprocal Debye length, $d\ln\epsilon/d\ln T = -1.34$ for water at room temperature and where $A(T)$ is the additive contribution from the solvation layer and the core material of the colloid. This expression is valid for arbitrary thickness of the double layer. The contribution stemming from the temperature dependence of surface charge to the Soret coefficient, which is given in the original work [37], is not considered here. This theoretical concept has been successfully applied in experimental studies [40] and was expanded for rod-like particles [39]

Lately it has been demonstrated that the thermophobicity of non-polar liquids is correlated with the ***heat of transport***, Q^* [41, 42]. The heat of transport, Q^*, is the heat carried by a unit diffusion flow of a component *1* when there is no temperature gradient and no diffusion of other components. The normal sign convention of thermodynamics is used, so that Q^* is positive if the system is heated and negative if the system loses heat. It turned out that the heat of transport in equimolar mixtures, which is a property of the pure components, is closely related to the Soret coefficient. Already in 1965 a similar correlation was observed for aqueous salt solutions [43].

Please note also that the thermal diffusion coefficient, D_{T} is is insensitive to a variation of the molecular weight M_w, while the ordinary diffusion coefficient scales with $D \propto M^{-\nu}$ with $\nu \approx 0.6$ for ideal chains. Experimental studies show this for polymers in organic solvents and water [44, 45, 46]. A theoretical explanation has been given by Brochard and de Gennes [47] stating that there is no hydrodynamic coupling between the monomers. More recent works show that deviations from the molar mass dependence are found for very short oligomers [48, 46].

One of the open questions is the size dependence of S_{T}, whether it is linear or quadratic [49]. While Duhr and Braun [34] observed an unambiguous quadratic dependence of S_{T} for carboxyl modified polystyrene (PS) beads in 1 mM TRIS buffer of different radii in the range from 20 nm to 1000 nm using a microscopic fluorescence technique, studies by Putnam and Cahill of carboxyl functionalized PS spheres in a size range from 26 nm to 92 nm (PS) gave some indication that the behavior could also be linear. Later Vigolo et al. [50] obtained a linear dependence investigating bis(2-ethylhexyl)sulfosuccinate (AOT) / isooctane / water microemulsion droplets with a radius between 1.8 nm to 16 nm. Unfortunately the shape of the microemulsion droplets was not properly characterized, and thus it is not certain that the microemulsion droplets are spherical in the range of study. Simultaneously the decreasing surface charge density of the AOT system with increasing radius will change the electrostatic contribution to the thermodiffusion properties. Recently, Braibanti et al. [51] repeated the experiment of Duhr and Braun [34] and studied the thermodiffusive behavior of highly diluted carboxyl modified PS spheres under the same conditions except that they used a 1:1 mixture of H_2O+D_2O to minimize sedimentation effects which can occur for the larger colloids. In the investigated radial range between 11 nm and 253 nm they found a linear size dependence for S_{T}. Recent studies of well-defined microemulsions of the type $\mathrm{H_2O}/n - \mathrm{alkane}/\mathrm{C_{12}E_5}$ (pentaethylene glycol monododecyl ether) using different n-alkanes support also a linear radial dependence [52, 53]. Also in theory the radial dependence is still controversially discussed. While Würger [54] expects a linear dependence for charged and uncharged solid colloids and only for soft colloids

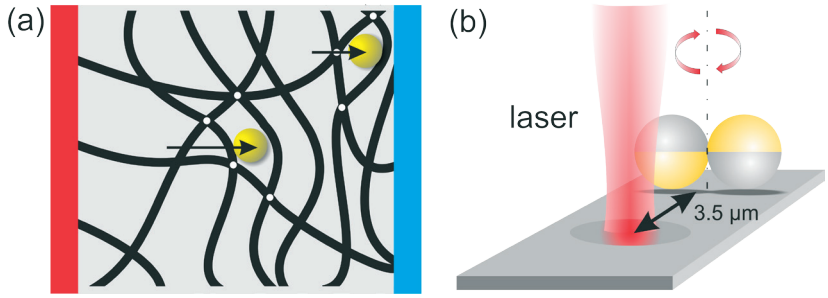


Fig. 4: *(a) Deformation of a transient polymer network with embeeded GNPs. The gray background represents the solvent continuum. The arrows show the local displacement in the presence of a temperature gradient.(b) Schematic illustration of a twin Janus particle rotating due to heating with a focused laser beam.*

a quadratic size dependence of S_{T}, Dhont *et al.* [37] predict also for charged solid colloids a quadratic dependence.
In conclusion we can state that the thermal diffusion in liquids and other soft matter systems is not understood, but the effect is very sensitive to changes in the interfacial layer between the solute and the surrounding solvent. Therefore it is also interesting for monitoring biochemical reactions, because those are often accompanied by changes in the hydration layer.

3 Interesting materials designed for applications in temperature gradients

In the following three subsections we will briefly summarize the results for interesting systems driven by a temperature gradient with a potential for practical applications.

3.1 Gold particles in polymer networks

An application of temperature gradients is the deformation of transient polymer networks with embedded gold nanoparticles (GNPs). Schwaiger and Köhler [55] studied GNPs in a highly entangled solution of an ultrahigh polymer (polystyrene $M_w = 16800$ kg/mol in toluene). As illustrated in figure 3 the GNPs are trapped in the polymer network, because the disentanglement or reptation time is very long. Heating of an isolated GNP by a focused laser beam leads to a local temperature gradient around the bead. Due to thermophoretic forces the solvent is attracted to the heated GNP, while the polymer is pushed away from the surface of the particle. Due to the entanglement of the polymers the transient polymer network is deformed. The movement and deformation of the network can be followed by optical microscopy.
It turned out that the displacement of the tracers and, hence, of the transient network is long-ranged and independent of distance in an infinite medium. Even particles far away from the local perturbation moved approximately by the same amount as particles in the immediate neighborhood of the heated GNP. In finite systems the displacement field is still far-reaching, but it decays to zero for distances of the order of the system size. In their experiments they have been able to observe significant displacements in a distance of 100 μm, corresponding to 10^3 colloid radii.

So far there is only this single experiment but in the future it will be interesting to perform more systematic experiments to study the influence of the molar mass and concentration on the network topology, disentanglement time and reversibility. With high speed cameras it will also be possible to study the elastic deformation and to perform micro rheology experiments in synthetic polymer networks or dense biological systems such as cells.

3.2 Janus Particles

Jiang *et al.* [56] studied the motion of half-metal coated silica or polystyrene particles (diameter = 1 μm) under laser irradiation. The motion is caused by self-thermophoresis: i.e., absorption of a laser at the metal-coated side of the particle creating a local temperature gradient around the particles which in turn drives the particle by thermophoresis. They observed that the particles moves faster under laser radiation and the motion follows the polarity of the Janus particle. At short times the rotational diffusion and the trapping effect can be neglected, and the motion can be treated as Brownian motion with a fixed directed motion. At longer times a trapping effect occurs, which can be explained by the two-dimensional trapping effect due to strong scattering from the metal part of the particle.
In pure water the motion of the Janus particles is toward the silica side, in other words, toward the colder side in the local temperature gradient created around the particle. In a solution with the non-ionic surfactant Triton X-100 the particles move toward the warm side. Finally they constructed a micromachine by carefully selecting twin Janus particles and focused a laser beam at 3.5 μm distance from the tethered particle (c.f. figure 3). Increasing the laser power they could observe a rotation of the twin particle. Above a certain laser power threshold the rotation speed increased linearly. They suspect that the threshold is caused by some interaction between the twin particle and the glass surface. The experiment shows that the external field induces self-propulsion and shows a new way to control micromachines. The basic principle is similar to the Crookes radiometer [57], but the scale is much more reduced.

3.3 Is it possible to build thermophoretic machines?

Especially in the last years has been a large research activity in the development and optimization of thermoelectric generators, which convert heat flux into electric current by the Seebeck effect [58] to recover waste heat. It turned out that the systems are quite robust, but the operational efficiency is quite low. Therefore, researcher look for novel strategies to recover waste heat. In this context researcher also think about the use of thermophoretic machines. Two conceptually different machines, a microgear and a turbine, have been suggested by computer simulations in the group of Ripoll [59, 60].
The proposed microgear [59] has a solid structure and the surface consists of eigth sawteeth with a asymmetric shape (c.f. figure 3.2). The microgear is surrounded by the multiparticle collision solvent and confined inside a circular wall. The rigid gear can freely rotate and its outer surface consists of a mono layer beads. The radial temperature field leads to a larger temperature gradient $\nabla T_{\mathrm{short}}$ along the short edge of the teeth compared to ∇T_{long} along the long edge. The bead solvent interaction is described by a Lennard Jones potential. The surface of the gear is in contact with the solvent, which is exposed to a temperature gradient. This leads to a thermophoretic force parallel to the edges of the gear. Depending on the interactions the thermophoretic forces can be along (thermophilic) or against (thermophobic) to the temperature

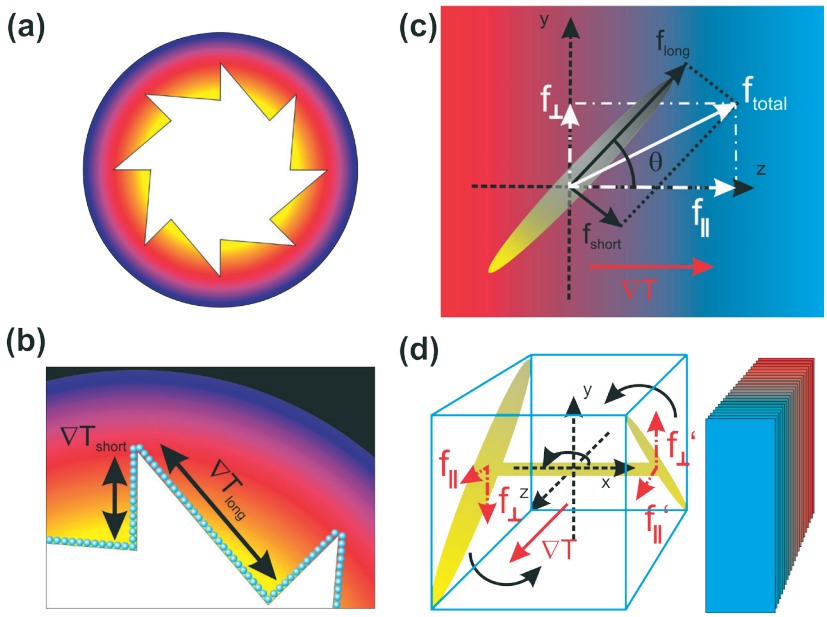


Fig. 5: *(a) Schematic illustration of an eight-teeth micro gear with a steady state radial temperature distribution. (b) Illustration of the temperature gradients* $\nabla T_{\mathrm{short}}$ *and* ∇T_{long} *along the short and long edges of the gear. (c) Diagram of the thermophoretic forces on an anisotropic ellipsoid. (d) Sketch of a minimalistic thermophoretic turbine. Two ellipsoidal blades have opposite orientation angles and their centers are connected by a rigid bond. The rotational direction of the turbine is parallel to the external thermal gradient.*

gradient. Due to the asymmetric gear geometry, this will lead to a non-vanishing torque on the gear, which leads to a unidirectional rotation.
The disadvantage of the microgear is that the temperature field has to be circular coupled to the gear. An interesting alternative is a thermophoretic turbine [60], which consists out of two ellipsoidal blades, which have opposite orientation angles and are coupled by a rigid bond (c.f. Fig. 3.2d). Figure 3.2 illustrates the thermophoretic forces acting on a simple ellipsoidal particle in a non-isothermal solution. The thermophoretic anisotropy of the particle is characterized by thermodiffusion factors along the short and long axes, so that the total thermophoretic force is the $f_{total} = f_{short} + f_{long}$, which can have a non-zero component, $f_\perp$ perpendicular to the thermal gradient, ∇T, which can create a net torque in the direction of the thermal gradient. Combining two ellipsoidal particles with opposite orientations connected by a rigid bond as illustrated in figure 3.2 we can construct a thermophoretic turbine. This turbine has a two-fold rotational symmetry with respect to the temperature gradient, but lacks reflection symmetry. It has been shown that the thermophoretic forces on the two ellipsoidal blades satisfy the relations, $f'_\parallel = f_\parallel$ and $f'_\perp = -f_\perp$, so that the torque acting on the turbine is parallel to the temperature gradient. The big advantage to the microgear is that the turbine is passively driven by an external temperature gradient, which does not have to be matched to the turbine. Therefore the turbine could be used to convert wasted heat into mechanical work.

4 Commercial Instruments utilizing temperature gradients

There are only two commercial instruments available utilizing temperature gradients for characterization of soft matter. The so-called thermal field flow fractionation method and the Microscale Thermophoresis. Both methods use thermophoresis for characterization and separation. Once the temperature gradients are know it is possible to the determine quantitatively the thermal transport and separation coefficients, but none of the methods has been validated versus the classical methods, yet. In the following we will briefly explain both methods including the experimental considerations.

4.1 Field-flow fractionation

The thermal field-flow fractionation (TFFF) belongs to family of flexible elution techniques capable of simultaneous separation and measurement. Besides temperature fields also other fields such as flow-, electric- or gravitational fields are used. The TFFF mechanism combines elements of chromatography and an external temperature field and can be used to characterize polymers or colloidal particles. The fundamental principle of TFFF is illustrated in Fig. 6. The separation of a sample takes place inside a narrow ribbon-like channel. This channel is composed of a thin piece of sheet material (usually 70-300 μm thick Mylar or polyimide) in which a channel is cut and which is usually clamped between two walls of highly-polished plane parallel surfaces through which a force can be applied. Fig. 6B illustrates the separation process. Component Y forms a distribution closest to the accumulation wall and is entrained in the lowest laminar flow band. It is gradually separated from component X moving in a faster band of the flow profile. How close the solute particles are to the accumulation wall is determined by the interplay between the applied field and homogeneity-restoring back diffusion. The retention time t_r is given approximately by

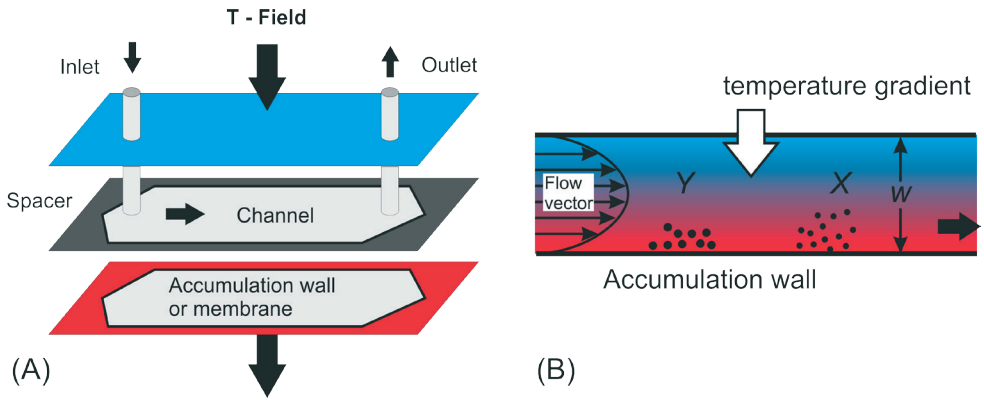


Fig. 6: *(A) Sketch of the basic components in most FFF channels. (B) Distribution of two arbitrary components and the unequal flow displacement velocities for the two species.*

$$\frac{t_r}{t^0} = \frac{|F|\, w}{6kT} \tag{8}$$

t^0 is the void time (the emergence time of a non-retained tracer), w is the channel thickness and $|F = kT(D/D_{\mathrm{T}})\Delta T/w|$ the strength of the thermal force depending on D and D_{T} the mutal and the thermal diffusion coefficient, respectively. A separation of two components is only possible when the force increment between the two species is large enough.

Experimental considerations Practically, thermal FFF (th-FFF) is used for the determination of molecular weights and molecular weight distributions. The channel is usually composed of two metallic blocks (with high thermal conductivity, preferably copper) with highly polished even surfaces between which a spacer ($\approx 100\mu$m) is clamped. The typical dimensions of the metal block are 40-60 cm length, 3-6 cm width and a thickness of 2-3 cm. The temperature gradient is applied perpendicular to the solvent flow. Usually the upper plate is heated to avoid convection. Often rather high temperature gradients, exceeding 10^6 K/m are applied, corresponding to temperature differences between the hot and the cold wall up to 100 K. The thermal diffusion drives the solute particles either more towards the cold or to warm side with a molecular weight or particle-size dependent penetration length of the concentration distribution into the parabolic channel flow profile, which then leads to a separation of the different molecular weight species. In the literature one often finds the statement that the macromolecules in a temperature gradient are always accumulate in the cold [61, 62, 24], but this is not always the case [63, 25].
Thermal FFF is especially suitable for synthetic polymers in organic solvent in the molecular range between $M = 10^4 - 10^7$ g/mol using temperature differences $\Delta T \sim 10 - 100$ K. The list of polymers which have been successfully characterized is as long as for other methods. The references to the original works can be found in the recent review by Cölfen and Antonietti [24]. While polymers in the range $M = 10^4 - 10^7$ g/mol are well resolved by th-FFF, polymers of lower molecular weight ($\sim 10^3$ g/mol) need an inconveniently high ΔT ($\sim$150 K) for retention, and problems of boiling solvent, etc. arise. However, successful separations of polystyrene down to 600 g/mol have been described using very high temperature gradients in a pressurized th-FFF channel [64].

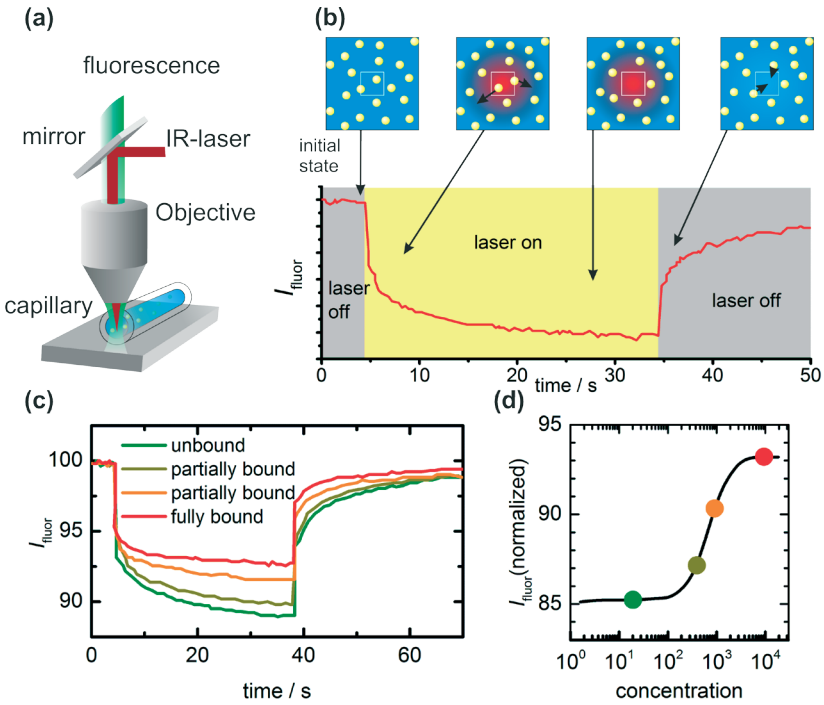


Fig. 7: *(a) Sketch of the MicroScale set-up (b) Temporal dependence of the fluorescent light intensity, when a heating laser is switched on and and off, when the fluorescently labelled particles accumulate in the cold region. (c) Schematic temporal fluorescent intensity scans for different concentrations of the substrate molecule. (d) Normalized fluorescent intensities as function of concentration. Further explainations see in the text.*

4.2 MicroScale Thermophoresis

The Microscale Thermophoresis (MST) [65, 66, 10] utilizes the thermophoresis effect to monitor biochemical reactions. The device consist of an epifluorescence microscope with an IR-laser (1480 nm), which is used to generate a temperature gradient due to strong absorption by water inside a capillary containing a solution of fluorescent molecules. While the temperature gradient is applied, the fluorophores in the solution are excited and the fluorescent light is observed through the same objective as the infrared laser (c.f. fig7a). Note that the IR-laser heats the sample, while the movement of the molecules is detected via fluorescence. The change in the fluorescent intensity is monitored as function of time, when the IR-laser is switched on and off (c.f. Fig. 7a). In the sketched cartoon the fluorescently labelled particles are thermophobic, so that the fluorescent intensity decreases towards a plateau value as function of time. If the IR-laser is switched off the system recovers (except for bleaching effects) back to the original fluorescent intensity. Figure 7c shows a typical binding experiment. When another type of molecule e.g. ligand is added in the solution, which can bind to the labeled molecules e.g. protein, we observe that the thermophoretic response of the latter is changed. This leads then to a change of the thermophoretic plateau, which depends on the fraction of molecules which have bound to the labeled molecule. In the displayed case the fluorescent plateau (*laser-on* status) increases monotonically till it reaches a plateau as function of concentration as sketched in figure 7d. This functional relation can then be used to calculate the equilibrium binding constant.

Experimental considerations The typical applied temperature gradients are in the order of 10^5 K/m. The sensitivity of the instrument is quite high and equilibrium binding constants in the pico molar range can be determined [10]. In general fluorescent labeling of the molecules is required. On one hand the fluorescent tag influence the diffusion of the solute molecules and on the other hand the coupling of a protein to the tag may alter or even inhibit the biochemical reaction [67]. Therefore, also an intrinsic fluorescence as in tryptophan is used [68] to investigate protein-ligand binding.
If a fluorescent label is used, there will be a low molar mass limit of the solute molecules. To minimize the influence on the diffusion the mass of the molecule with the fluorescent tag should be at least 10-times larger than the mass of the fluorescent label. Typical fluorescent labels have masses in the order of 350 g/mol, therefore the solute molecules should have at least a mass of 35 kg/mol. There is no upper molar mass limit and also large colloids with a diameter of 1 μm can be investigated [34].
For a quantitative analysis of the thermophoretic quantities such as S_{T} or D_{T} a precise temperature measurement is required. Typically this is done by measuring the fluorescent intensity of a temperature dependent dye [69]. This method is not very accurate because often artifacts occur due to adsorption and bleaching effects. Additionally a proper thermometer calibration is required. Nevertheless for the characterization of biochemical reactions MST is very useful device, especially due to its low sample consumption.

5 Outlook

Thermal diffusion or thermophoresis is one of the unsolved problems in statistical physics and requires still a lot of experimental and theoretical work. Nevertheless in the recent years many interesting experiments have been performed showing that is possible to manipulate colloids

due to temperature gradients. Here is especially the optical creation of a temperature gradient an important tool, because the method is quite flexible and does not require the fabrication of special microfluidic instruments. Additionally the development of the microscale thermophoresis instrument turned out to be extremely useful in biotechnolgy. Additionally, thermophoresis provides an alternative strategy to design smart micromachines and has become a promising tool in the field of microfluidics and conversion of waste heat. As many of the developments took place only in the last years, we can expect interesting future developments.

Appendices

A Theoretical description of thermal diffusion

There is presently no microscopic theory available to describe the thermal diffusion and thermophoresis phenomena. Nowadays are still the phenomenological equations used, which were derived by Onsager [70, 71] .
In irreversible processes which are not so far from equilibrium, a system can be divided into small subsystems and local equilibrium is assumed [72]. That is, every "small" volume element δV fulfills microscopic reversibility. Furthermore, the amount of particles inside δV is sufficient to define thermodynamic properties, such as entropy and temperature. In the whole system, the macroscopic irreversible processes are assumed to obey the same laws as the average regressions of fluctuations in the microscopic equilibrium systems. In order to achieve local equilibrium, several conditions should be satisfied [72]: *first*, the thermodynamic forces should be sufficiently small to hold the linear response of the field conjugated flux; *second*, the whole system should be sufficiently close to equilibrium; *third*, the characteristic distances over which the thermodynamic forces vary should be sufficiently large so that these forces can be viewed as being constant over the microscopic length scale required to properly define a local thermodynamic state; *fourth*, the characteristic times over which the thermodynamic forces vary should be sufficiently long that these forces can be viewed as being constant over the microscopic times required to properly define a local thermodynamic state. A single transport process in linear response to an imposed field, where the conjugated flux is proportional to the corresponding thermodynamic force, can be described by equations proposed more than one and a half centuries ago [73]:

$$\vec{J} = -L\vec{X} \tag{9}$$

where $\vec{J}$ is the flux, $\vec{X}$ the thermodynamic force, and L a proportionality constant, or transport coefficient. For example, Fouriers law describes the first relation of such a type [74]:

$$\vec{J_q} = -\lambda\vec{\nabla}T \tag{10}$$

Here the heat flux $\vec{J_q}$ is related to a temperature gradient $\vec{\nabla}T$ with the proportionality constant λ, which is termed the "thermal conductivity" and characterizes the capability of the material to transport heat through the system. Ficks first law [74] is another example for the flux of matter $\vec{J_m}$ (the amount passing through unit area in unit time) caused by the molar concentration gradient $\vec{\nabla}c_m$:

$$\vec{J_m} = -D\vec{\nabla}c_m \tag{11}$$

where D is the diffusion coefficient, which has the dimension square length per time in SI-units ($m^2\ s^{-1}$). In an incompressible binary system regarding only the translational diffusion

the system is characterized by one single diffusion coefficient, which is commonly notated as "mutual diffusion coefficient" D_{12}. Note that the value for D_{12} can be equally assigned in different frames of reference (e.g. mass-fixed, mole-fixed, ...) [74]. In the following discussion, we will use the "mass-fixed frame of reference", where no net transfer of total mass crosses a fixed reference plane, to connect Onsagers relation with Ficks law.
Two or more transport processes may happen simultaneously and interfere with each other. Onsager reciprocal relations [70, 71] describe such coupling between irreversible transport processes, where the macroscopic flux is expressed as a linear combination of different forces. Therefore, the coupling processes of heat conduction and matter transfer in a multi-component mixture can be written in the following way.

$$\begin{aligned} \vec{J}_i &= \sum_{j=1}^{n-1} L_{ij}\vec{X}_j + L_{iq}\vec{X}_q, \quad i = 1, ..., n-1 \\ \vec{J}_q &= \sum_{j=1}^{n-1} L_{qj}\vec{X}_j + L_{qq}\vec{X}_q \end{aligned} \tag{12}$$

where n denotes the total number of species, $\vec{X}_j$ the field (thermodynamic force) which drives the mass flux of species j, $\vec{X}_q$ the field which drives the heat flux, and $L_{\alpha\beta}$ the Onsager coefficients with $\alpha, \beta = (j, q)$. We will introduce their specific representations below. Note that this representation assumes that the flux is parallel to the driving force, otherwise each $L_{\alpha\beta}$ has to be a 3×3 tensor. In this work, only binary isotropic media are studied, where the driving force is always parallel to the flux. Hence, without loss of generality, we can use the scalar form to represent the Onsager relation in a binary system:

$$\begin{aligned} J_1 &= L_{11}X_1 + L_{1q}X_q \\ J_q &= L_{q1}X_1 + L_{qq}X_q \end{aligned} \tag{13}$$

where the chemical-potential gradient and the temperature gradient are the thermodynamic forces [73]:

$$X_1 = -\nabla_T \left[(\mu_1 - \mu_2)\right]/T \tag{14}$$

$$X_q = \nabla T/T^2 \tag{15}$$

Here μ_i is the chemical potential of species i (note the chemical potential here is defined in (J/kg) instead of (J/mol) as the "mass-fixed frame of reference" is used).$\nabla_T \mu$ denotes a chemical potential gradient at constant temperature.
Of the two diagonal elements in equation 13, L_{11} is related to diffusion under a concentration gradient; and L_{qq} is connected to heat conduction under a temperature gradient. Of the two off-diagonal elements, L_{1q} is related to thermal diffusion, the diffusion of mass under a temperature gradient and L_{q1} is related to the Dufour effect, the heat transport under a concentration gradient. The assumption of microscopic reversibility of the system requires a general reciprocal relation to hold for equilibrium fluctuations as well as the macroscopic irreversible processes which obeys the same laws of fluctuation regression [4, 5]. According to the reciprocal relation, the two off-diagonal elements are equal: $L_{1q} = L_{q1}$. In the following discussion, we shall link the Onsager coefficients to the transport coefficients shown in equations 10 and 11.
First, if we assume the steady state of the non-equilibrium system is reached where the mass flux is zero, the relation between the thermal conductivity and Onsager coefficients can be derived

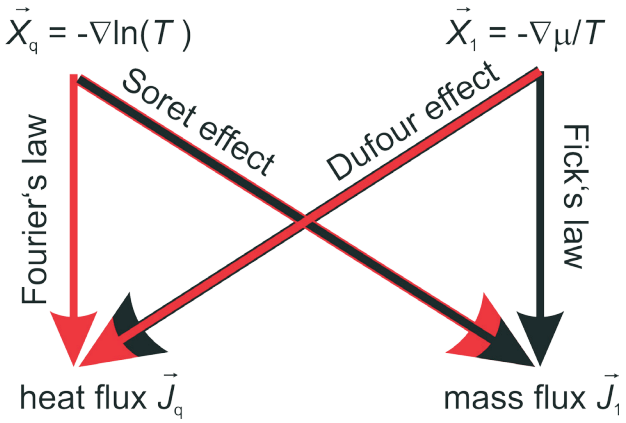


Fig. 8: *Illustration of the cross effects in a binary mixture in a temperature gradient. The generalized force $\vec{X}_q$ causes a heat flux and additionally a mass flux due to the Soret effect, while the generalized force $\vec{X}_1$ causes a mass flux as well as a heat flux by the Dufour effect. Due to the high heat conduction in fluids compared to gases the Dufour effect can experimentally not be observed in fluids.*

from equations 10, 13 and 15:

$$\lambda_{J_1=0} = \frac{-1}{T^2}\left[L_{qq} - \frac{L_{1q}^2}{L_{11}}\right] = -\left(\frac{J_q}{\nabla T}\right)_{J_1=0} \tag{16}$$

When the temperature gradient disappears, the mass diffusion of the system is only caused by the chemical potential gradient, and from equation 13 follows,

$$J_1|_{\nabla T=0} = -L_{11}\frac{\nabla_T\left(\mu_1 - \mu_2\right)}{T} \tag{17}$$

If the "mass fixed frame of reference" is used in equation 11, and the unit of the mutual diffusion coefficient is kept as "m^2s^{-1}", the molar concentration gradient may be replaced by the dimensionless weight fraction of species 1, c_1, and equation 11 can be rewritten as:

$$J_1|_{\nabla T=0} = -\rho D_{12}\nabla c_1 \tag{18}$$

where ρ is the overall mean density of the system given in "kg/m^3". It can be assumed to be a constant property since the system is close to equilibrium. Combining the two equations, we obtain the expression for the mutual diffusion coefficient as:

$$D_{12} = L_{11}\frac{\nabla_T\left(\mu_1 - \mu_2\right)}{\rho T\nabla c_1} \tag{19}$$

According to the Gibbs-Duhem relation [73], at constant temperature and pressure, the chemical potential of a multicomponent system has the following relation:

$$\sum_i w_i d\mu_i = 0 \tag{20}$$

Note that we also use dimensionless weight fraction of species i, w_i, to replace the normally used particle number to be consistent with the special definition of chemical potential here. Hence, in a binary system, the chemical potential gradient can be expressed as:

$$\nabla_T \left(\mu_1 - \mu_2\right) = \frac{1}{\left(1 - c_1\right)} \frac{\partial \mu_1}{\partial c_1} \nabla c_1 \tag{21}$$

Combining equations 19 and 21, the relation between D_{12} and L_{11} is written as:

$$D_{12} = \frac{L_{11}}{\rho \left(1 - c_1\right) T} \frac{\partial \mu_1}{\partial c_1} = \frac{L_{11}}{\rho T} \frac{1}{c_1 \left(1 - c_1\right)} \left(\frac{\partial \ln c_1}{\partial \mu_1} \right)^{-1} \tag{22}$$

The remaining question is how to link the transport coefficients for thermal diffusion with the Onsager coefficients. Phenomenologically, the law of thermal diffusion is defined as [75]:

$$J_1 = -D_{12} \rho \nabla c_1 - D_{\mathrm{T}} \rho c_1 \left(1 - c_1\right) \nabla T \tag{23}$$

Comparing to equation 13, we obtain the link between D_{T} and L_{1q}:

$$D_{\mathrm{T}} = \frac{1}{\rho T^2} \frac{1}{c_1 \left(1 - c_1\right)} L_{1q} \tag{24}$$

Finally, the definition of Soret coefficient is obtained for a system in the steady state, where $J_1 = -J_2 = 0$:

$$S_{\mathrm{T}} = \frac{D_{\mathrm{T}}}{D} = -\frac{1}{c_1 \left(1 - c_1\right)} \frac{\nabla c_1}{\nabla T} \tag{25}$$

References

[1] C. Ludwig. Diffusion zwischen ungleich erwärmten orten gleich zusammengesetzter lösungen. *Sitz. ber. Akad. Wiss. Wien Math.-naturw. Kl*, 20:539, 1856.

[2] C. Soret. Sur l'etat d'quilibre que prend au point de vue de sa concentration une dissolution saaline primitivement homogne dont deux parties sont portes a des tempratures diffrentes. *Arch. Geneve*, 3:48–64, 1879.

[3] J. Jeans. *An Introduction to the Kinetic Theory of Gases*. At the University Press, Cambridge, 1952.

[4] J. Janek, C. Korte, and A. B. Lidiard. Thermodiffusion in ionic solids - model experiments and theory. *Lect. Notes. Phys.*, 584:146–183, 2002.

[5] F. Montel, J. Bickert, A. Lagisquet, and G. Galliero. Initial state of petroleum reservoirs: A comprehensive approach. *J. Petrol. Sci. Eng.*, 58:391–402, 2007.

[6] A. Bahat, S. R. Caplan, and M. Eisenbach. Thermotaxis of human sperm cells in extraordinarily shallow temperature gradients over a wide range. *Plos One*, 7, 2012.

[7] K. Okabe, N. Inada, C. Gota, Y. Harada, T. Funatsu, and S. Uchiyama. Intracellular temperature mapping with a fluorescent polymeric thermometer and fluorescence lifetime imaging microscopy. *Nat. Commun.*, 3:705–1–705–, 2012.

[8] Guillaume Baffou, Herve Rigneault, Didier Marguet, and Ludovic Jullien. A critique of methods for temperature imaging in single cells. *Nat Meth*, 11:899–901, 2014.

[9] S. K. R. Williams and D. Lee. Field-flow fractionation of proteins, polysaccharides, synthetic polymers, and supramolecular assemblies. *Journal of Separation Science*, 29:1720–1732, 2006.

[10] M. Jerabek-Willemsen, T. Andr, W. Wanner, H.M. Roth, S. Duhr, P. Baaske, and D. Breitsprecher. Microscale thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.*, pages 101–113, 2014.

[11] C. B. Mast, N. Osterman, and D. Braun. Could thermal gradients drive molecular evolution? *Curr. Org. Chem.*, 17:1732–1737, 2013.

[12] C. B. Mast and D. Braun. Thermal trap for dna replication. *Phys. Rev. Lett.*, 104:188102–1–188102–4, 2010.

[13] G. Chen, M. S. Dresselhaus, G. Dresselhaus, J. P. Fleurial, and T. Caillat. Recent developments in thermoelectric materials. *Int. Mater. Rev.*, 48:45–66, 2003.

[14] E. Y. Buchin and Y. I. Denisenko. Use of thermomigration in mems technology - art. no. 626011. *P. Soc. Photo-Opt. Ins.*, 6260:L2601–L2601, 2006.

[15] C. Chen, H. Y. Hsiao, Y. W. Chang, F. Y. Ouyang, and K. N. Tu. Thermomigration in solder joints. *Mater. Sci. Eng. R*, 73:85–100, 2012.

[16] K. Clusius and G Dickel. Neues Verfahren zur Gasentmischung und Isotopentrennung. *Naturwissenschaften*, 26:546, 1938.

[17] K. Clusius and G Dickel. Zur Trennung der Chlorisotope. *Naturwissenschaften*, 27:148–149, 1939.

[18] L. Waldmann. *Transporterscheinungen in Gasen von mittleren Druck*, volume 12, pages 295–514. Springer Verlag, Berlin, 1958.

[19] S. Chapman. The kinetic theory of simple and composite monatomic gases viscosity, thermal conduction, and diffusion. *Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character*, 93:1–20, 1916.

[20] R. Fürth. An elementary theory of thermal diffusion. *P. Roy. S. - Math. Phy.*, 179:0461–0469, 1942.

[21] L. Waldmann. Die Temperaturerscheinungen bei der Diffusion in ruhenden Gasen und ihre messtechnische Anwendung, journal = Zeitschrift für Physik, year = 1947, volume = 124, pages = 2-30.

[22] J. J. Hurly and M. R. Moldover. Ab initio values of the thermophysical properties of helium as standards. *J. Res. Natl. Inst. Stan.*, 105:667–688, 2000.

[23] S. Wiegand. Thermal diffusion in liquid mixtures and polymer solutions. *J. Phys.: Condens. Matter*, 16:R357–R379, 2004.

[24] H. Cölfen and M. Antonietti. *Field-Flow Fractionation Techniques for Polymer and Colloid Analysis*, volume 150 of *Advances in Polymer Science*, pages 67–187. 2000.

[25] B. J. de Gans, R. Kita, S. Wiegand, and J. Luettmer-Strathmann. To the hot or to the cold? Unusual thermal diffusion in polymer solutions. *Phys. Rev. Lett.*, 91:245501, 2003.

[26] S. Hartmann, G. Wittko, W. Köhler, K. I. Morozov, K. Albers, and G. Sadowski. Thermophobicity of liquids: Heats of transport in mixtures as pure component properties. *Phys. Rev. Lett.*, 109:065901–1–065901–4, 2012.

[27] W. M. Rutherford. Effect of mass-distribution on the isotopic thermal-diffusion of substituted benzenes. *J. Chem. Phys.*, 81:6136–6139, 1984.

[28] C. Debuschewitz and W. Köhler. Molecular origin of thermal diffusion in benzene plus cyclohexane mixtures. *Phys. Rev. Lett.*, 87:055901, 2001.

[29] G. Wittko and W. Köhler. Universal isotope effect in thermal diffusion of mixtures containing cyclohexane and cyclohexane-d(12). *J. Chem. Phys.*, 123, 2005.

[30] P. Polyakov, J. Luettmer-Strathmann, and S. Wiegand. Study of the thermal diffusion behavior of alkane/benzene mixtures by thermal diffusion forced rayleigh scattering experiments and lattice model calculations. *J. Phys. Chem. B*, 110:26215–26224, 2006.

[31] S. Iacopini, R. Rusconi, and R. Piazza. The ”macromolecular tourist”: Universal temperature dependence of thermal diffusion in aqueous colloidal suspensions. *Eur. Phys. J. E*, 19:59–67, 2006.

[32] Y. Kishikawa, S. Wiegand, and R. Kita. Temperature dependence of soret coefficient in aqueous and nonaqueous solutions of pullulan. *Biomacromolecules*, 11:740–747, 2010.

[33] Z. Wang, H. Kriegs, and S. Wiegand. Thermal diffusion of nucleotides. *J. Phys. Chem. B*, 116:7463–7469, 2012.

[34] S. Duhr and D. Braun. Thermophoretic depletion follows boltzmann distribution. *Phys. Rev. Lett.*, 96:168301–1–168301–4, 2006.

[35] R. D. Astumian. The unreasonable effectiveness of equilibrium theory for interpreting nonequilibrium experiments. *Am. J. Phys.*, 74:683–688, 2006.

[36] R. Sugaya, B. A. Wolf, and R. Kita. Thermal diffusion of dextran in aqueous solutions in the absence and the presence of urea. *Biomacromolecules*, 7:435–440, 2006.

[37] J. K. G. Dhont, S. Wiegand, S. Duhr, and D. Braun. Thermodiffusion of charged colloids: Single-particle diffusion. *Langmuir*, 23:1674–1683, 2007.

[38] J. K. G. Dhont and W. J. Briels. Single-particle thermal diffusion of charged colloids: Double-layer theory in a temperature gradient. *Eur. Phys. J. E*, 25:61–76, 2008.

[39] Z. Wang, H. Kriegs, J. Buitenhuis, J. K. G. Dhont, and S. Wiegand. Thermophoresis of charged colloidal rods. *Soft Matter*, 9:8697–8704, 2013.

[40] H. Ning, J. K. G. Dhont, and S. Wiegand. Thermal-diffusive behavior of a dilute solution of charged colloids. *Langmuir*, 24:2426–2432, 2008.

[41] S. Hartmann, W. Köhler, and K. I. Morozov. The isotope soret effect in molecular liquids: A quantum effect at room temperatures. *Soft Matter*, 8:1355–1360, 2012.

[42] S. Hartmann, G. Wittko, F. Schock, W. Gross, F. Lindner, W. Kohler, and K. I. Morozov. Thermophobicity of liquids: Heats of transport in mixtures as pure component properties-the case of arbitrary concentration. *J. Chem. Phys.*, 141, 2014.

[43] B. D. Butler and J. C. R. Turner. Soret coefficients and heats of transport of some aqueous electrolytes at 9 degrees. *J. Phys. Chem.*, 69:3598–3599, 1965.

[44] J. C. Giddings, K. D. Caldwell, and M. N. Myers. Thermal-diffusion of polystyrene in 8 solvents by an improved thermal field-flow fractionation methodology. *Macromolecules*, 9:106–112, 1976.

[45] J. Rauch and W. Köhler. Diffusion and thermal diffusion of semidilute to concentrated solutions of polystyrene in toluene in the vicinity of the glass transition. *Phys. Rev. Lett.*, 88:185901–1–185901–4, 2002.

[46] Z. Wang, D. Afanasenkau, M. Dong, D. Huang, and S. Wiegand. Molar mass and temperature dependence of the thermodiffusion of polyethylene oxide in water/ethanol mixtures. *J. Chem. Phys.*, 141:064904–1–064904–7, 2014.

[47] F. Brochard and P. G. Degennes. Soret effect of flexible macromolecules. *C.R. Acad. Sci. II*, 293:1025–1027, 1981.

[48] J. Rauch and W. Köhler. On the molar mass dependence of the thermal diffusion coefficient of polymer solutions. *Macromolecules*, 38:3571–3573, 2005.

[49] R. Piazza and A. Parola. Thermophoresis in colloidal suspensions. *J. Phys-Condens. Mat.*, 20:153102, 2008.

[50] D. Vigolo, G. Brambilla, and R. Piazza. Thermophoresis of microemulsion droplets: Size dependence of the soret effect. *Phys. Rev. E*, 75:040401, 2007.

[51] M. Braibanti, D. Vigolo, and R. Piazza. Does thermophoretic mobility depend on particle size? *Phys. Rev. Lett.*, 100:108303–1–108303–4, 2008.

[52] Ph. Naumann, N. Becker, S. Datta, T. Sottmann, and S. Wiegand. Soret coefficient in nonionic microemulsions: Concentration and structure dependence. *J. Phys. Chem. B*, 117:5614–5622, 2013.

[53] Ph. Naumann, S. Datta, T. Sottmann, B. Arlt, H. Frielinghaus, and S. Wiegand. Isothermal behavior of the soret effect in nonionic microemulsions: Size variation by using different n-alkanes. *J. Phys. Chem. B*, 118:3451–3460, 2014.

[54] A. Würger. Thermal non-equilibrium transport in colloids. *Reports on Progress in Physics*, 73:126601, 2010.

[55] F. Schwaiger and W. Köhler. Photothermal deformation of a transient polymer network. *Macromolecules*, 46:1673–1677, 2013.

[56] H.-R. Jiang, N. Yoshinaga, and M. Sano. Active motion of a janus particle by self-thermophoresis in a defocused laser beam. *Phys. Rev. Lett.*, 105:268302–1268302–4, 2010.

[57] R. Piazza. 'thermal forces': Colloids in temperature gradients. *J. Phys.-Condens. Mat.*, 16:S4195–S4211, 2004.

[58] F. J. DiSalvo. Thermoelectric cooling and power generation. *Science*, 285:703–706, 1999.

[59] M. C. Yang and M. Ripoll. A self-propelled thermophoretic microgear. *Soft Matter*, 10:1006–1011, 2014.

[60] M.C. Yang, R. Liu, M. Ripoll, and K. Chen. A microscale thermophoretic turbine driven by external diffusive heat flux. *Nanoscale*, 6:13550–13554, 2014.

[61] J. C. Giddings, M. Martin, and M. N. Myers. Thermo-gravitational field-flow fractionation - elution thermo-gravitational column. *Separ. Sci. Technol.*, 14:611–643, 1979.

[62] M. Martin and R. Reynaud. Polymer analysis by thermal field-flow fractionation. *Anal. Chem.*, 52:2293–2298, 1980.

[63] M. Giglio and A. Vendramini. Soret-Type Motion of Macromolecules in Solution. *Phys. Rev. Lett.*, 38(1):26–30, 1977.

[64] J. C. Giddings, L. K. Smith, and M. N. Myers. Thermal field-flow fractionation - extension to lower molecular-weight separations by increasing liquid temperature-range using a pressurized system. *Anal. Chem.*, 47:2389–2394, 1975.

[65] C. J. Wienken, P. Baaske, U. Rothbauer, D. Braun, and S. Duhr. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.*, 1:1–7, 2010.

[66] F. M. Weinert, C. B. Mast, and D. Braun. Optical fluid and biomolecule transport with thermal fields. *Phys. Chem. Chem. Phys.*, 13:9918–9928, 2011.

[67] M. M. Baksh, A. K. Kussrow, M. Mileni, M. G. Finn, and D. J. Bornhop. Label-free quantification of membrane-ligand interactions using backscattering interferometry. *Nat. Biotechnol.*, 29:357–U173, 2011.

[68] S. A. I. Seidel, C. J. Wienken, S. Geissler, M. Jerabek-Willemsen, S. Duhr, A. Reiter, D. Trauner, D. Braun, and P. Baaske. Label-free microscale thermophoresis discriminates sites and affinity of protein-ligand binding. *Angew. Chem.-Int. Ed.*, 51:10656–10659, 2012.

[69] S. Duhr, S. Arduini, and D. Braun. Thermophoresis of dna determined by microfluidic fluorescence. *Eur. Phys. J. E*, 15:277–286, 2004.

[70] L. Onsager. Reciprocal Relations in Irreversible Processes. I. *Phys. Rev.*, 37:405–426, 1931.

[71] L. Onsager. Reciprocal Relations in Irreversible Processes. II. *Phys. Rev.*, 38:2265–2279, 1931.

[72] D.J. Evans and G.P. Morriss. *Statistical Mechanics of Nonequilibrium Liquids*. London, 1990.

[73] S.R. de Groot and P. Mazur. *Non-Equilibrium Thermodynamics*. Dover, New York, 1984.

[74] E.L. Cussler. *Diffusion: Mass Transfer in Fluid Systems*. Cambridge University Press, 1984.

[75] K.E. Grew and T.L. Ibbs. *Thermal Diffusion in Gases*. University Press, Cambridge, 1952.

F 5 Microgels

W. Richtering
Institute of Physical Chemistry
RWTH Aachen University

Contents

1 Introduction 2

2 Temperature sensitive microgels 3

3 Core-shell microgels 5

3.1 Seed-and-feed synthesis of core-shell microgels 5
3.2 Multifunctional core-shell microgels 6
3.3 Polyelectrolyte binding to core-shell microgels 7

4 Hollow temperature sensitive microgels 8

Acknowledgement 10

References 11

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

A microgel is a special form of a macromolecule that combines features from different architectures; this is illustrated in. Like a gel it is a chemically cross-linked polymer swollen in a good solvent. However the spatial dimensions are much smaller which is why the are termed “microgel“. Often the size is smaller than 1 µm and sometimes such gels are called “nanogels“.

The presence of cross-links distinguishes microgels from (hyper-)branched polymers that contain branching points but no cross-links. Cross-linking provides the topological integrity and thus microgels differ from colloidal supramolecular aggregates.The surface of a microgel is characterized by dangling chains that reach into the solution and thus a microgels resembles a sterically stabilized latex particle.

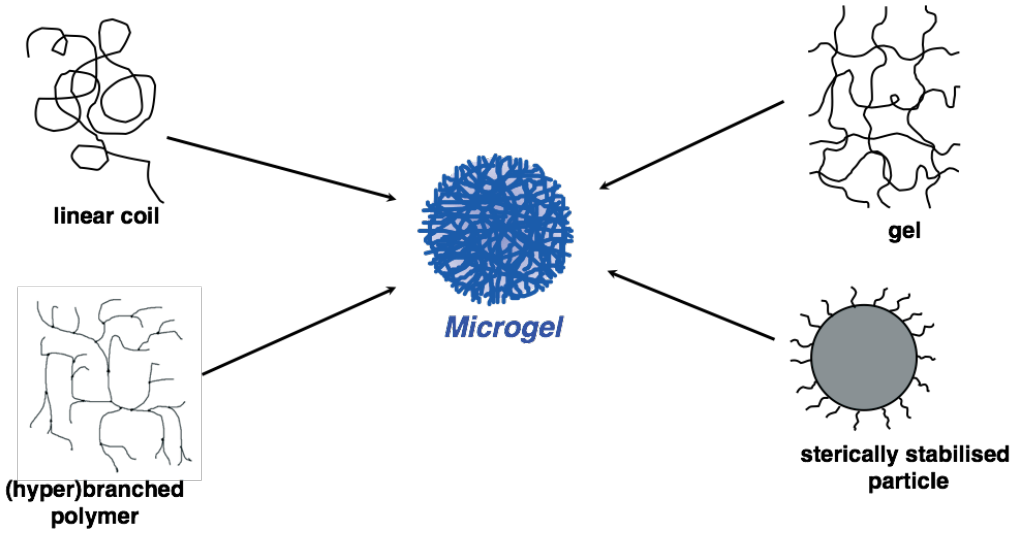


Fig. 1: *Illustration of the macromolecular architecture of a microgel.*

Microgel dispersions gained great interest as promising candidates for various applications as, e.g., in the printing and pharmaceutical industries as well as for drug delivery, separation, catalysis, and microoptics. Microgels are also interesting model systems in the research area of soft condensed matter, where they allow the investigation of the structure and dynamics of concentrated colloidal suspensions. Since the particle interaction forces can be controlled by the properties of the particle, microgels are ideal systems to study the relation between the interaction potential, phase behavior and flow properties.

Microgels can be prepared via different routes, most often the applied techniques employ radical polymerization of monomers and cross-linker either in miniemulsion[i] or as precipitation polymerization.[ii] An important alternative is the cross-linking of prepolymers.[iii]

2 Temperature sensitive microgels

The solubility of many aqueous polymer systems changes upon variation of the temperature. The most widely studied water-swellable, temperature-sensitive microgel system is based on poly-*N*-isopropylacrylamide (PNiPAM) first prepared by Pelton et al.[iv] This polymer shows phase separation upon heating at ca 32°C. The chemical structures of the monomer and of a cross-linker that is often employed are shown in together with the phase diagram.[v]

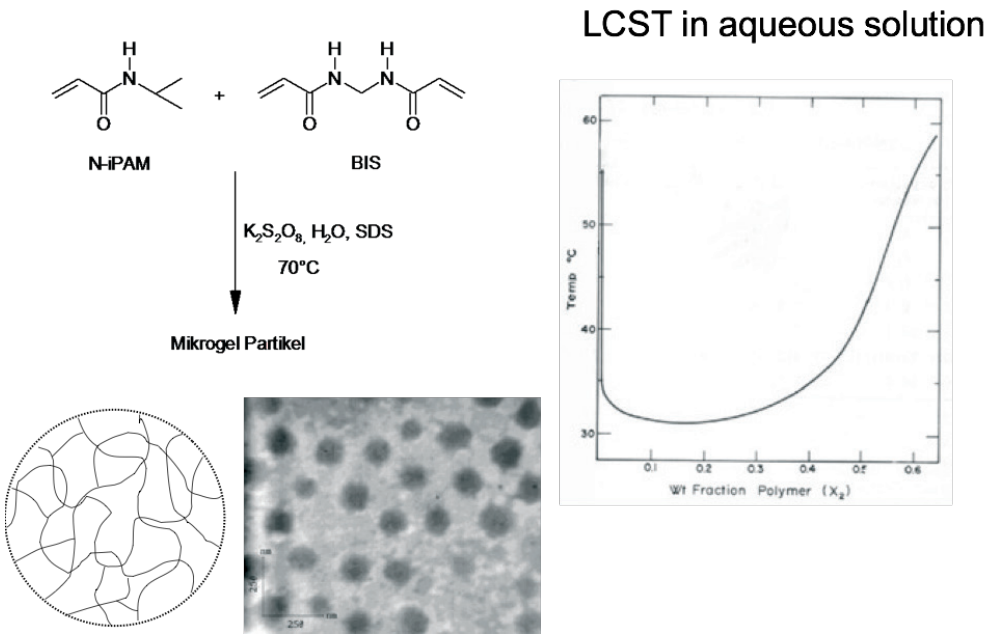


Fig. 2: *Chemical structures of NiPAM and cross-linker and the phase diagram of PNiPAM in water.*

Typically the polymerization kinetics of cross-linker and monomer are not identical.[vi] This can lead to an inhomoheneous distribution of cross-linker inside the microgel particle, especially during precipitation polymerization. Often the cross-linking density decreases from the center of the particle to the surface[vii , viii] but that can be controlled via the polymerisation conditions.[ix]

Scattering methods including small-angle neutron scattering (SANS) , small-angle x-ray scattering (SAXS) as well as static and dynamic light scattering (SLS, DLS) are well suited to investigate in detail the structure of PNiPAM microgels. From static scattering experiments, a form factor can be obtained which describes the structure of microgel particles accounting for both the inhomogeneous internal network structure and the overall particle shape. Depending on the particle size one needs to employ light, neutron or X-ray scattering or a combination of them. Basically these methods provide similar information; however, the experimental range of momentum transfer q and the contrast are different. Thus a suitable choice has to be made depending on microgel properties. Often small angle neutron scattering (SANS) is most suitable, however, deuterated solvents or mixtures of deuterated and hydrogenated solvents are used in SANS and one should keep in mind that phase boundaries are often slightly

different between H- and D-containing solvents due to the changes in the hydrogen bonding to water.

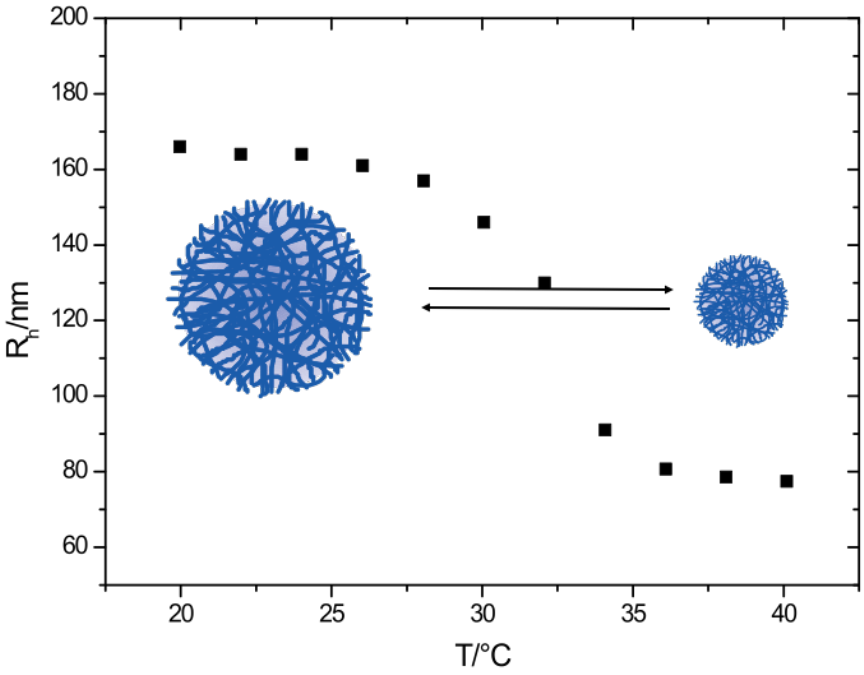


Fig. 3: *Temperature dependent size (as determined from the diffusion coefficient via dynamic light scattering) of PNiPAM microgels in water.*

shows how the size of PNiPAM microgel particles decreases sharply upon heating, above the volume phase transition temperature (VPTT). Dynamic light scattering is probably the most frequently used technique to determine the size of microgel particles. DLS detects the collective diffusion coefficient and the Stokes-Einstein equation connects the diffusion coefficient at infinite dilution with the hydrodynamic radius R_h. Microgels are strongly swollen by the solvent and usually the non-draining limit is reached, i.e. the solvent is immobilized inside the microgel and the hydrodynamic radius that is influenced by the dangling chains at the particle surface.

The size difference between swollen and collapsed state depends on the cross-linker content of the microgels and can thus be controlled via the monomer feed during the polymerization. The colloidal stability of the microgels at temperatures above the VPTT also depends on the polymerization conditions. Flocculation is e.g. observed when a surfactant has been present during precipitation polymerization but has been removed afterwards.

3 Core-shell microgels

The presence of irreversible chemical cross-links allows preparing microgels with compartmentalized internal structure.

3.1 Seed-and-feed synthesis of core-shell microgels

In particular core-shell microgels can be prepared, typically in a two-step seed and feed polymerisation.[x] illustrates this procedure that cen be used to prepare microgels with double temperature sensitivity. Here two polymers with different VPTTs are empolyed: poly-*N*-isopropylacrylamide (PNiPAM) and poly-*N*-isopropylmethacrylamide (PNiPMAM). [xi]

A feed and seed polymerization leads to different particles as compared to a random copolymerization. In the latter case one can also obtain microgels containing different temperature sensitive moieties that can also reveal peculiar behaviours[xii] and e.g. an internally phase separated structure.[xiii]

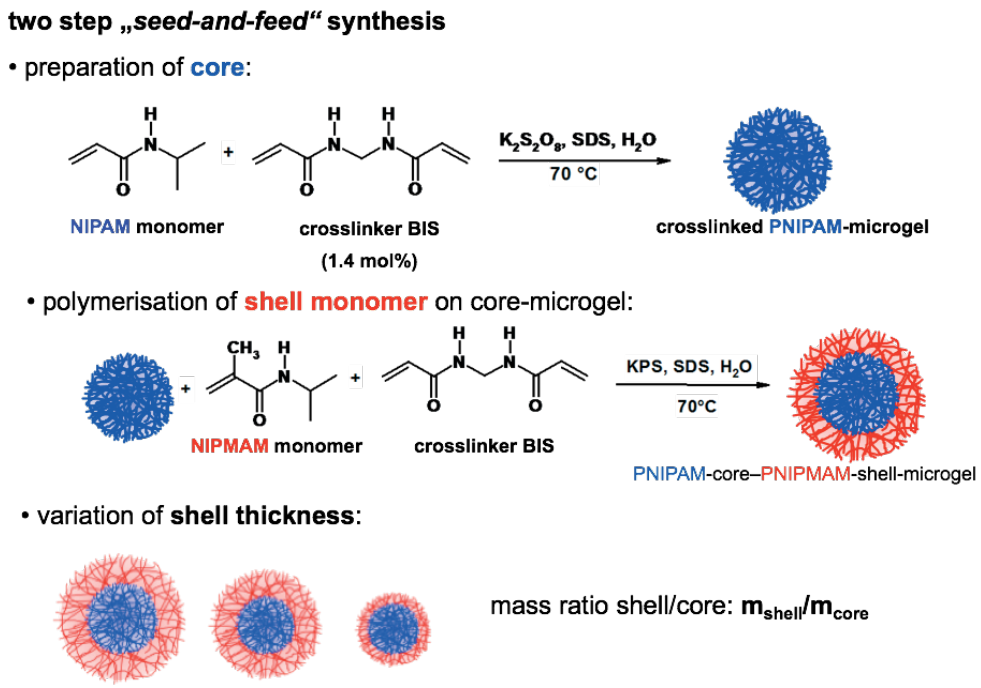


Fig. 4: *Seed-and-feed synthesis of core-shell microgels based on NiPAM and NIPMAM monomers.*

3.2 Multifunctional core-shell microgels

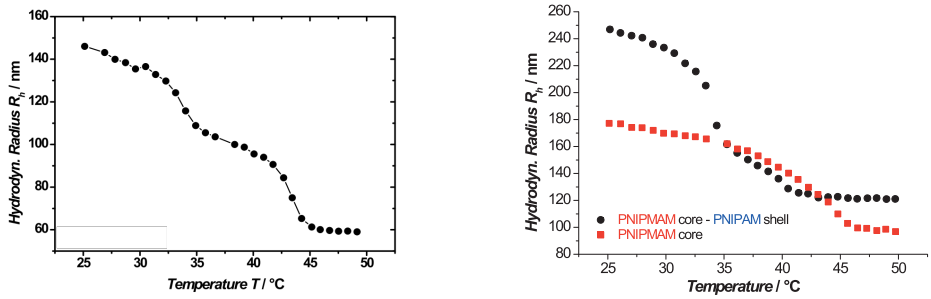


Fig. 5: *Temperature dependent size of core-shell microgels. Left: PNiPAM core and PNiPMAM shell; right PNiPAM core and PNiPAM shell.*

Such core-shell thermo-responsive particles reveal complex swelling transitions with temperature, as shown in. Microgels with PNiPAM core and PNiPMAM shell (left) reveal a clear two step swelling ipon heating: first the core swell and then the shell swells. However, the swelling properties of such core-shell microgels can be altered via cross-link densities and mass ration of core and shell polymers.[xiv]

The behavior of the inverse system, i.e. with PNiPMAM core and PNiPAM shell is even more complex, see the right plot in: at intermediate temperature, the collapsed shell restrains the swelling of the core: a corsett effect is observed. The internal structure of these microgels has been determined in detail by means of SANS. [xv,xvi]

Core-shell systems can be also prepared with solid cores that provide a rigid substrate and thus strongly influence the swelling of the shell.[xvii] Such systems are of great interest as hybrid materials[xviii], however, the size and shape of the rigid part will not be affected by the swelling of the network.

Multisensitive temperature and pH- resposnive core-shell microgels can be obtained if acidic and/or basic functional groups are incorporated. The swelling properties depend on the presence and spatial distribution of counterions. Here the complex internal structure of microgels plays an inportant role: the counterions penetrate into the microgels and lead to an enhanced osmotic pressure. Many different temperature and pH-sensitive core-shell microgels have been reported in the literature. [x,xix,xx xxi, xxii]

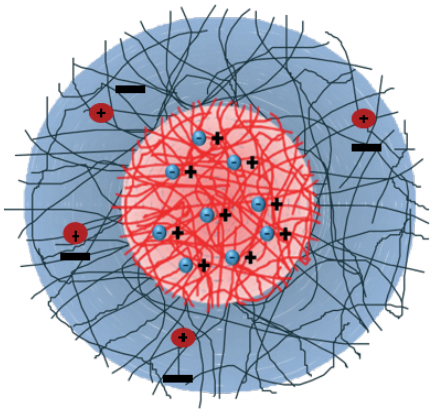


Fig. 6: *Core-shell microgel functionalized with different charged groups in core and shell, respectively. The counterions are shown as well.*

3.3 Polyelectrolyte binding to core-shell microgels

Charged core-shell are an important class of soft matter with many applications. It is important to remember that both regions, i.e. core and shell are chemically cross-linked. That leads to (i) a clear compartmentalisation and (ii) to mechanical coupling of the both regions as discussed above. It also makes core-shell gels distinctly different from (hyper-) branched polymers, where one can also have different chemical functionalities near the centre of the branched molecules and at the end of the branches. However, in branched polymers, backfolding of the chains can occur, which is not possible in core-shell networks. Thus microgels allow preparing compartmentalized particles. These particles are highly swollen and thus allow for the penetration not only by the solvent but also by molecules that are sufficiently small as compared to the mesh size of the microgel network. This allows for uptake and release of guest molecules. As one example we discuss the binding of oppositely charged polyelectrolytes to charged core-shell microgels.

reveals the comlex binding of polycations to a negatively charged microgel where the charged groups are located in the core of the microgel. The zeta potential of the microgels is very small although the microgel core contains acid groups. This is due to the fact the conterions are mostly inside the microgel and the zeta potential is caused by the small number of charged moieties of initiator fragments present near the surface of the swollen microgel.

These microgels are able to bind oppositely charged polyelectrolytes. Short chain polycations can fully penetrate into the microgels and thus this binding process does not lead to charge reversal. On the other hand, long chain polycations cannot fully penetrate leading to charge reversal upon binding.[xxiii]

The binding of polyelectrolytes to multiresponsive microgels is rather complex[xxiv ,xxv,xxvi] and there are many more reports in the literature than given here. Charged microgels can also be emplyed for the layer-by-layer depostion of polyelectrolytes.[xxvii, xxviii,xxii]

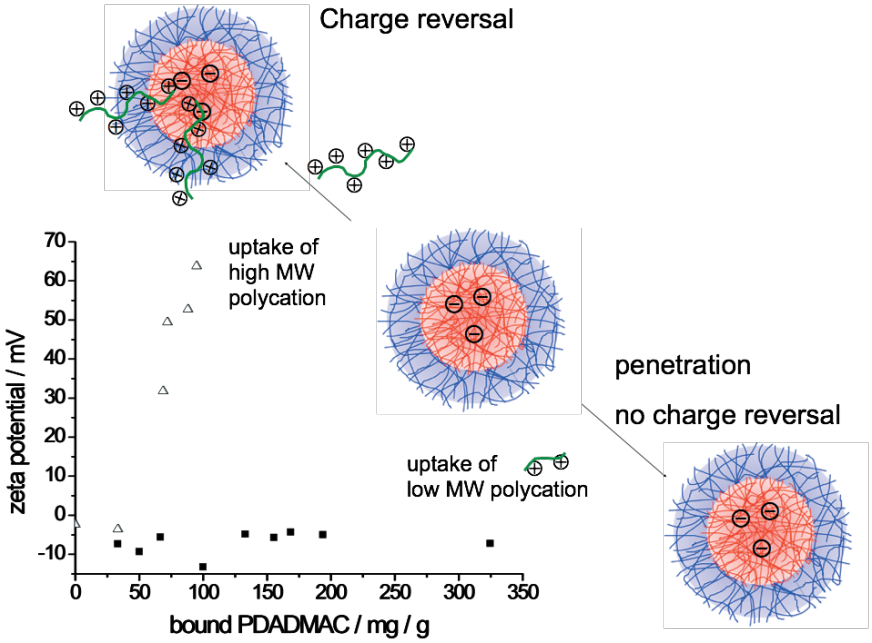


Fig. 7: *Binding of oppositely charged polyelectrolytes of different molar mass to a charged core-shell microgel.*[xxiii] *Please note that the low molar mass counterions are not shown.*

4 Hollow temperature sensitive microgels

Finally we want to mention shortly a rather new field: hollow responsive microgels. Hollow capsules are in the focus of the rational design of e.g. responsive drug delivery systems that store and protect drugs from degradation in vivo but allow drug release e.g. in response to an external stimulus.

Nanocapsules with permeable chemically cross-linked shell can be prepared by encapsulating a sacrificial core by a shell of a cross-linked swollen polymer network. Afterwards the core is dissolved, and it is typically assumed that the microgels are hollow with the void size being given by the size of the sacrificial core as illustrated in

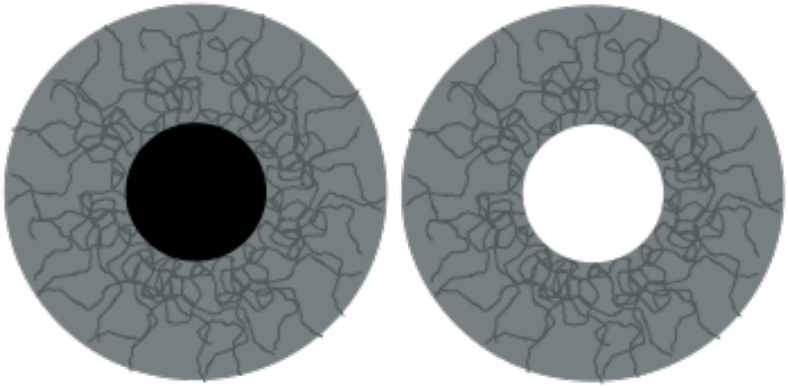


Fig. 8: *Schematic drawing of: Left: Hybrid core-shell microgel with rigid sacrificial core; right: hollow microgels after core removal.*

hollow nanoparticles made of temperature-sensitive polymers like poly-N-isopropylacrylamide (PNIPAM) are particularly promising materials and have been prepared based on this approach.[xxix,xxx,xxxi] The temperature-dependent void size and shell thickness is crucial for the rational design of new, functional drug delivery vehicles and can be determined by means of SANS. Here the contrast variation technique is very important.

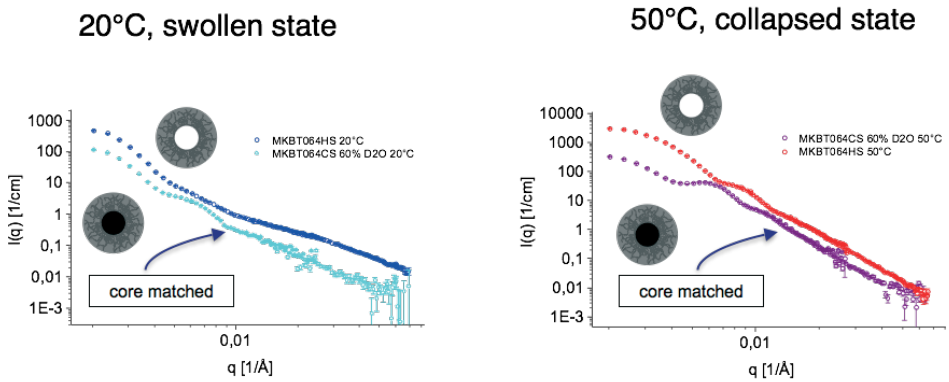


Fig. 9: *SANS curves from core-shell and hollow microgels at different temperatures.*

Fig. 9 shows scattering data from SANS experiments with core-shell and hollow microgels.[xxxii] The experiments with the core shell particles were performed such that the scattering contrast of the silica core was matched. Thus scattering intensity of both systems – core-shell and hollow microgel – is dominated by the polymer shell. It is obvious that the structure of the shell in the hollow microgel is different as compared to the shell in the core-shell particle. The swelling properties of the shell depend on its cross-link density and mass and a structure-sensitivity dilemma of responsive hollow nanogels has been discussed. [xxxii]
Very recently hollow microgels with two temperature sensitive shells have been resported as well. [xxxiii]

Acknowledgement

We thank the Deutsche Forschungsgemeinschaft for funding within the SFB 985 Functional microgels and microgel systems.

References

[i] Landfester, K.; Musyanovych, A. Hydrogels in Miniemulsions. *Adv Polym Sci* **2010**, *234*, 39–63.

[ii] Pich, A.; Richtering, W. Microgels by Precipitation Polymerization: Synthesis, Characterization, and Functionalization. *Adv Polym Sci* **2010**, *234*, 1–37.

[iii] Albrecht, K.; Moeller, M.; Groll, J. Nano- and Microgels Through Addition Reactions of Functional Oligomers and Polymers. *Adv Polym Sci* **2010**, *234*, 65–93.

[iv] Pelton, R. H.; Chibante, P. Preparation of Aqueous Latices with N-Isopropylacrylamide. *Colloids Surf. A: Physicochem. Eng. Asp.* **1986**, *20*, 247–256.

[v] Heskins, M.; Guillet, J. E. Solution Properties of Poly(N- Isopropylacrylamide) . *J. Macromol. Sci. A Chem.* **1968**, *2*, 1441–1455.

[vi] Wu, X.; Pelton, R. H.; Hamjelec, A. E.; Woods, D. R.; W, M. The Kinetics of Poly(N-Isopropylacrylamide) Microgel Latex Formation. *Colloid Polym Sci* **1994**, *272*, 467–477.

[vii] Guillermo, A.; Cohen Addad, J. P.; Bazile, J. P.; Duracher, D.; Elaissari, A.; Pichot, C. NMR Investigations Into Heterogeneous Structures of Thermosensitive Microgel Particles. *J Polymer Sci Part B: Polym Phys* **2000**, *38*, 889–898.

[viii] Stieger, M.; Richtering, W.; Pedersen, J. S.; Lindner, P. Small-Angle Neutron Scattering Study of Structural Changes in Temperature Sensitive Microgel Colloids. *J Chem Phys* **2004**, *120*, 6197–6206.

[ix] Meyer, S.; Richtering, W. Influence of Polymerization Conditions on the Structure of Temperature-Sensitive Poly(N-Isopropylacrylamide) Microgels. *Macromolecules* **2005**, *38*, 1517–1519.

[x] Jones, C. D.; Lyon, L. A. Synthesis and Characterisation of Multiresponsive Core-Shell Microgels. *Macromolecules* **2000**, *33*, 8301–8306.

[xi] Berndt, I.; Richtering, W. Doubly Temperature Sensitive Core-Shell Microgels. *Macromolecules* **2003**, *36*, 8780–8785.

[xii] Keerl, M.; Richtering, W. Synergistic Depression of Volume Phase Transition Temperature in Copolymer Microgels. *Colloid Polym Sci* **2007**, *285*, 471–474.

[xiii] Keerl, M.; Pedersen, J. S.; Richtering, W. Temperature Sensitive Copolymer Microgels with Nanophase Separated Structure. *J. Am. Chem. Soc.* **2009**, *131*, 3093–3097.

[xiv] Berndt, I.; Pedersen, J. S.; Lindner, P.; Richtering, W. Influence of Shell Thickness and Cross-Link Density on the Structure of Temperature-Sensitive Poly-N-Isopropylacrylamide-Poly-N-Isopropylmethacrylamide Core-Shell Microgels Investigated by Small-Angle Neutron Scattering. *Langmuir* **2006**, *22*, 459–468.

[xv] Berndt, L.; Pedersen, J. S.; Richtering, W. Structure of Multiresponsive “Intelligent” Core-Shell Microgels. *J. Am. Chem. Soc.* **2005**, *127*, 9372–9373.

[xvi] Berndt, I.; Pedersen, J. S.; Richtering, W. Temperature-Sensitive Core–Shell Microgel Particles with Dense Shell. *Angew. Chem. Int. Ed.* **2006**, *45*, 1737–1737.

[xvii] J. H. Kim and M. Ballauff, *Colloid Polym Sci*, 1999, **277**, 1210–1214.

[xviii] M. Karg, *Colloid Polym. Sci.*, 2012, **290**, 673.

[xix] William H. Blackburn, L. Andrew Lyon, *Colloid Polym Sci*. **2008**, 286, 563-569

[xx] Gan, D.; Lyon, L. A. Interfacial Nonradiative Energy Transfer in Responsive Core-Shell Hydrogel Nanoparticles. *J. Am. Chem. Soc.* **2001**, *123*, 8203–8209.

[xxi] Schachschal, S.; Balaceanu, A.; Melian, C.; Demco, D. E.; Eckert, T.; Richtering, W.; Pich, A. Polyampholyte Microgels with Anionic Core and Cationic Shell. *Macromolecules* **2010**, *43*, 4331–4339.
[xxii] Gelissen, A. P. H.; Schmid, A. J.; Plamper, F. A.; Pergushov, D. V.; Richtering, W. Quaternized Microgels as Soft Templates for Polyelectrolyte Layer-by-Layer Assemblies. *Polymer* **2014**, *55*, 1991–1999.
[xxiii] Kleinen, J.; Klee, A.; Richtering, W. Influence of Architecture on the Interaction of Negatively Charged Multisensitive Poly(N-Isopropylacrylamide)-Co-Methacrylic Acid Microgels with Oppositely Charged Polyelectrolyte: Absorption vs Adsorption. *Langmuir* **2010**, *26*, 11258–11265.
[xxiv] Ballauff, M. Spherical Polyelectrolyte Brushes. *Prog Polym Sci* *Prog Polym Sci* **2007**, *32*, 1135–1151.
[xxv] Hansson, P.; Bysell, H.; Månsson, R.; Malmsten, M. Peptide–Microgel Interactions in the Strong Coupling Regime. *J. Phys. Chem. B* **2012**, *116*, 10964–10975.
[xxvi] Kleinen, J.; Richtering, W. Rearrangements in and Release From Responsive Microgel– Polyelectrolyte Complexes Induced by Temperature and Time. *The Journal of Physical Chemistry B* **2011**, *115*, 3804–3810.
[xxvii] Wong, J. E.; Gaharwar, A. K.; Muller-Schulte, D.; Bahadur, D.; Richtering, W. Dual-Stimuli Responsive PNiPAM Microgel Achieved via Layer-by-Layer Assembly: Magnetic and Thermoresponsive. *J. Colloid Interface Sci.* **2008**, *324*, 47–54.
[xxviii] Wong, J. E.; Díez-Pascual, A. M.; Richtering, W. Layer-by-Layer Assembly of Polyelectrolyte Multilayers on Thermoresponsive P(NiPAM-Co-MAA) Microgel: Effect of Ionic Strength and Molecular Weight. *Macromolecules* **2009**, *42*, 1229–1238.
[xxix] Zha, L. S.; Zhang, Y.; Yang, W. L.; Fu, S. K. Monodisperse Temperature-Sensitive Microcontainers. *Adv Mater* **2002**, *14*, 1090.
[xxx] Nayak, S.; Gan, D.; Serpe, M. J.; Lyon, L. A. *Small* **2005**, *1*, 416.
[xxxi] Lapeyre, V.; Renaudie, N.; Dechezelles, J.-F.; Saadaoui, H.; Ravaine, S.; Ravaine, V. Multiresponsive Hybrid Microgels and Hollow Capsules with a Layered Structure. *Langmuir* **2009**, *25*, 4659–4667.
[xxxii] Dubbert, J.; Honold, T.; Pedersen, J. S.; Radulescu, A.; Drechsler, M.; Karg, M.; Richtering, W. How Hollow Are Thermoresponsive Hollow Nanogels? *Macromolecules* **2014**, *47*, 8700–8708.
[xxxiii] Dubbert, J.; Nothdurft, K.; Karg, M.; Richtering, W. Core-Shell-Shell and Hollow Double-Shell Microgels with Advanced Temperature Responsiveness. *Macromol. Rapid Commun.* **2015**, doi: 10.1002/marc.201400495

F 6 Drug delivery in blood

D. A. Fedosov, K. Müller, and G. Gompper
Institute of Complex Systems
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**
1.1 Blood . . . 2
1.2 Blood cells . . . 2
1.3 Drug delivery . . . 3

2 Methods and models **4**
2.1 Blood cells and suspended particles . . . 4
2.2 Simulation setup and definitions . . . 6

3 Results **7**
3.1 Particle margination . . . 7
3.2 Dependence of margination on particle size . . . 9
3.3 Dependence of margination on vessel size . . . 11
3.4 Dependence of margination on particle shape . . . 11

4 Discussion and conclusions **12**

A Dissipative particle dynamics **13**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

1.1 Blood

Blood is circulated around the entire body performing a number of physiological functions. Its main functions are the transport of oxygen and nutrients to cells of the body, removal of waste products such as carbon dioxide and urea, and circulation of molecules and cells which mediate the organism's defense and immune response and play a fundamental role in the tissue repair process. Abnormal blood flow is often correlated with a broad range of disorders and diseases which include, for instance, hypertension, anemia, atherosclerosis, malaria, and thrombosis. Understanding the rheological properties and dynamics of blood cells and blood flow is crucial for many biomedical and bioengineering applications. Examples include the development of blood substitutes, the design of blood flow assisting devices, and drug delivery. In addition, understanding of vital blood related processes in health and disease may aid in the development of new effective treatments.

Blood is a physiological fluid that consists of erythrocytes or red blood cells (RBCs), leukocytes or white blood cells (WBCs), thrombocytes or platelets, and plasma containing various molecules and ions. RBCs constitute about 45% of the total blood volume, WBCs around 0.7%, and the rest is taken up by blood plasma and its substances. One microliter of blood contains about 5×10^6 RBCs, roughly 5000 WBCs, and approximately 2.5×10^5 platelets.

1.2 Blood cells

Figure 1 shows a scanning electron micrograph of blood cells. Human RBCs have a relatively

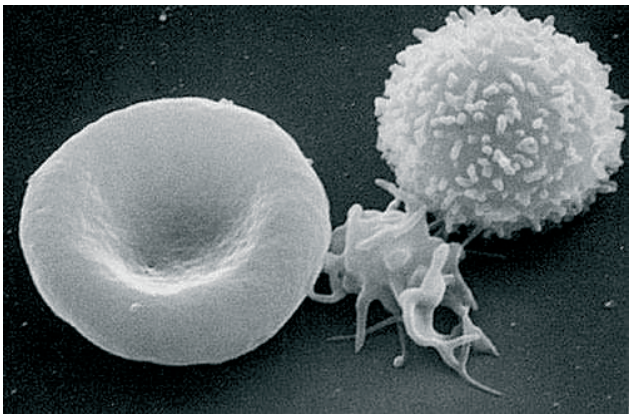


Fig. 1: *A scanning electron micrograph of blood cells. From left to right: human erythrocyte, thrombocyte (platelet), leukocyte.*

simple structure in comparison to other cells. RBCs resemble biconcave disks with an average diameter of approximately $8\,\mu$m and contain a viscous cytosol enclosed by a membrane. A RBC membrane consists of a lipid bilayer with an attached cytoskeleton formed by a network of the spectrin proteins linked by short filaments of actin. At the stage of the RBC formation, the nucleus and other organelles that are generally present in other eukaryotic cells are ejected, leaving behind a relatively homogeneous cytoplasm and no inner cytoskeleton. RBC cytoplasm

is a hemoglobin rich solution, which is able to bind oxygen. Therefore, the main RBC function is oxygen supply and delivery to body tissues. RBCs are extremely deformable and can pass through capillaries with a diameter several times smaller than the RBC size.
In comparison to RBCs, WBCs are spherical in shape with a diameter between $7\,\mu$m and $20\,\mu$m. WBCs have one or multiple nuclei and are stiffer than RBCs. However, WBCs are also able to undergo significant deformation when entering the smallest blood capillaries. WBCs are an important part of the body's immune system. They protect the body against invading bacteria, parasites, and viruses by killing these microorganisms through phagocytosis ingestion and other antigen-specific cytotoxic mechanisms. There exist different types of leukocytes (e.g., neutrophils, eosinophils, basophils, monocytes, and lymphocytes), each of which is designed to fight a specific type of infection. WBCs may adhere to the vascular endothelium , which is important for their physiological function in the immune response.

1.3 Drug delivery

The use of targeted micro- and nano-carriers for the delivery of imaging agents and drugs provides a promising strategy for early detection and treatment of diseases, e.g., of cancer [1, 2]. However, the design of particles carrying different contrast agents and drugs as well as their physical delivery are very challenging tasks. Micro- and nano-particle fabrication, which needs to address several issues such as bio-compatibility, durability, binding to specific targets, and the ability of controlled release, has been strongly advanced in recent years [3, 4, 5]. Nevertheless, the development of efficient strategies for the delivery of carriers, including their distribution in the organism following systemic administration [6] and their transport through biological barriers [6, 7, 8] (e.g., microvascular walls, interstitial space, and cell membranes), requires a much more detailed understanding of the relevant physical and biological mechanisms [2, 6, 9, 10].
Successful delivery of micro- and nano- carriers strongly depends on their efficient binding to specific targeted sites. Consequently, the distribution of carriers within vessel cross-sections plays an important role, since binding of carriers is only possible in case of direct particle-wall interactions. The cross-sectional distribution of micro- and nano-particles depends on several relevant parameters, which concern blood flow properties (such as flow rate, RBC deformability, and hematocrit – the volume fraction of RBCs), vessel size, and particle characteristics (such as size, shape, and deformability). The migration of various suspended particles or cells toward walls in blood flow, which is often referred to as *margination*, has been observed experimentally for white blood cells [11, 12], platelets [13], and rigid micro-particles [14, 15]. Particle margination is mediated by RBCs, which migrate to the vessel center due to hydrodynamic interactions with the walls (called lift force) [16, 17] leading to a RBC-free layer near the walls. More precisely, the occurrence of margination is a consequence of the competition between lift forces on RBCs and suspended particles, and their interactions in flow [18]. However, the dependence of margination efficiency on particle size and shape remains largely unexplored so far.
The role of particle size and shape in the efficient delivery is a multi-faceted problem. Large enough particles with a characteristic diameter (D_p) greater than about $4\,\mu$m may become trapped in the smallest capillaries of the body [19]. In addition, recent experiments suggest that large particles with $D_p \gtrsim 3\,\mu$m are subject to an enhanced phagocytosis [20]. However, recent microfluidic experiments [21] have shown that spheres with the size of $2\,\mu$m show a significantly higher adhesion density than particles with a size of $200\,$nm and $500\,$nm. Other experiments [22] indicate that liposomes with $D_p < 70\,$nm and $D_p > 300\,$nm have shorter

circulation times than those having an intermediate size of $D_p \approx 150 - 200$ nm. Furthermore, nano-particles with a size below $20 - 30$ nm are rapidly excreted through the kidneys [23]. Adhesion of different particles has been studied experimentally [24, 25] and theoretically [26, 27], with the result that oblate ellipsoids are subject to stronger adhesion than spheres with the same volume. To better understand the adhesion potential of micro- and nano-particles, a quantitative description of particle margination under realistic blood flow conditions is required.
In this chapter, we investigate the role of particle size and shape on the margination efficiency, and therefore on their adhesion potential. Several sizes ranging from about hundred nanometers to a few micrometers and two different shapes (circular and elliptical) are considered. The margination of micro- and nano-particles is studied numerically for a wide range of hematocrit values, vessel sizes, and flow rates using a two-dimensional (2D) model. Our results indicate that large particles possess a larger probability of being marginated than small particles. As the particle size becomes very small (less than about $100 - 200$ nm), the particle distribution within vessel cross-section can be described well by the plasma volume around flowing RBCs. Furthermore, circular particles marginate better than ellipses, however the adhesion efficiency of elliptical particles is expected to be superior in comparison to that of circles due to their smaller adhesion area.

2 Methods and models

To represent fluid flow, the dissipative particle dynamics (DPD) method [28, 29] is employed. DPD is a mesoscopic particle-based simulation approach which properly captures hydrodynamics often described by the incompressible Navier-Stokes equation. Detailed description of the DPD method can be found in Appendix A.

2.1 Blood cells and suspended particles

To study margination of micro- and nano-particles in blood flow, it is sufficient to model only RBCs, since volume fraction of the other blood cells is negligible in comparison to the volume fraction of RBCs. The RBC volume fraction is referred to as hematocrit H_t. In 2D, RBCs are modeled by a collection of $N_v = 50$ particles connected by $N_s = N_v$ springs [30] with the potential [31]

$$V_{spring} = \sum_{j \in 1 \dots N_s} \left[\frac{k_B T l_m (3x_j^2 - 2x_j^3)}{4p(1 - x_j)} + \frac{k_p}{l_j} \right], \tag{1}$$

where l_j is the length of the spring j, l_m is the maximum spring extension, $x_j = l_j/l_m$, p is the persistence length, $k_B T$ is the energy unit, and k_p is the spring constant. A balance between the two force terms in Eq. (1) determines a non-zero equilibrium spring length l_0. The cell model also incorporates a bending energy between two consecutive springs given by

$$V_{bend} = \sum_{j \in 1 \dots N_v} k_b \left[1 - \cos(\theta_j) \right], \tag{2}$$

where k_b is the bending constant and θ_j is the instantaneous angle between two adjacent springs having the common vertex j. In addition, a constraint to maintain a constant cell area is imposed on each cell by the potential

$$V_{area} = k_a \frac{(A - A_0)^2}{2A_0}, \tag{3}$$

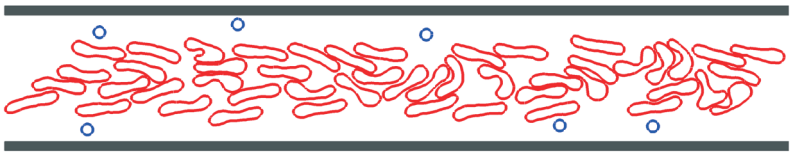


Fig. 2: *Snapshot of blood flow in a micro-channel in 2D. RBCs are colored in red and suspended particles in blue. Simulation conditions correspond to* $H_t = 0.3$ *and* $\dot{\gamma}^* \approx 29.3$.

where k_a is the area constraint coefficient, A is the instantaneous RBC area, and A_0 is the specified (target) area.

The RBC diameter is chosen to be $D_r \equiv L_0/\pi$, where $L_0 = N_s l_0$ is the cell perimeter. For comparison in physical units, typical value for healthy RBCs is $D_r = 6.1$ μm. The combination of A_0 and L_0 determines the shape of a RBC which is characterized by the reduced area $A^* = 4A_0/(\pi D_r^2) = 0.46$. The RBC equilibrium spring length is $l_0 = 0.063 D_r$ and $l_m/l_0 = 2.2$. In all simulations $k_b = 50\ k_B T$ for RBCs. The other important non-dimensional number, which characterizes the ratio of RBC elasticity to bending rigidity, is $\alpha = Y D_r^2/\kappa$, where $Y = (-\partial^2 V_{spring}/\partial l^2)|_{l_0}$ is the stretching modulus and $\kappa = k_b l_0$ is the RBC bending rigidity. $\alpha = 1340$ for RBCs [30], which has been roughly estimated by mimicking RBC stretching experiment [32]. The area constraint coefficient is set to $k_a = 37210 k_B T/D_r^2$.

Micro- and nano-particles are modeled by a collection of N_v^p particles, which are constrained to maintain a rigid configuration. Two shapes of particles are used in simulations including circular and elliptical particles. The particles are characterized by a size D_p, which corresponds to the diameter for circular particles and to the longest axis for elliptical particles.

Coupling between the fluid flow and cells/carriers is achieved through viscous friction [31] between cell vertices and the surrounding fluid particles, which is implemented via the DPD interactions F^D and F^R, see Appendix A. The strength γ of the dissipative force F^D for the interaction between a fluid particle and a membrane vertex is computed such that no-slip BCs are ensured. The derivation of γ is based on the idealized case of linear shear flow over a flat boundary with length L. The total shear force exerted by the fluid on the length L is equal to $L\eta\dot{\gamma}$, where η is the fluid's viscosity and $\dot{\gamma}$ is the local wall shear-rate. The same fluid force has to be also transmitted onto a discrete membrane having N_L vertices within the length L. The force on a single membrane vertex exerted by the sheared fluid can be found as $F_s = \int_{A_h} n g(r) F^D dA$ where n is the fluid number density, $g(r)$ is the radial distribution function of fluid particles with respect to the membrane particles, and A_h is the half circle volume of fluid above the membrane. Here, the total shear force on the length L is equal to $N_L F_s$. The equality of $N_L F_s = L\eta\dot{\gamma}$ results in an expression of the dissipative force coefficient in terms of the fluid density and viscosity, wall density N_L/L, and r_c. Under the assumption of linear shear flow the shear rate $\dot{\gamma}$ cancels out. This formulation results in satisfaction of the no-slip BCs for the linear shear flow over a flat membrane; however, it also serves as an excellent approximation for no-slip at the membrane surface. Note that conservative interactions between fluid and membrane particles are turned off, which implies that the radial distribution function is structureless, $g(r) = 1$.

size	$D_p = 0.3D_r$	$D_p = 0.63D_r$	$D_p = 0.15D_r$	$D_p = 0.04D_r$
N_v^p	20	20	15	10
$N_{W=20\,\mu m}$	6	6	20	100
$N_{W=10\,\mu m}$	6	/	14	/
$N_{W=40\,\mu m}$	12	/	30	/

Table 1: *Carrier characteristics. N_v^p is the number of particles per carrier and N is the number of carriers in the system depending on the channel width and particle size.*

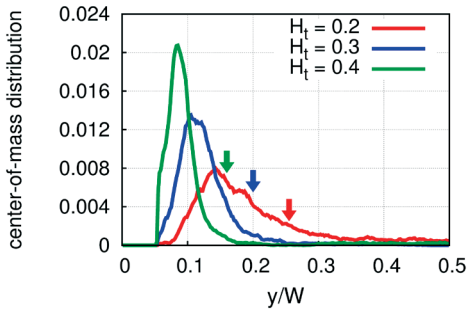


Fig. 3: *Center-of-mass distributions of carriers for various H_t values at $\dot{\gamma}^* \approx 29.3$. Simulation results for circular particles with $D_p = 0.3D_r$ (1.83 μm). The wall is at $y/W = 0$. The arrows indicate the boundary of the RBCFL for the different hematocrits, marked by corresponding colors.*

2.2 Simulation setup and definitions

The simulation setup consists of a single slit-like channel with different widths $W = 10$, 20, and 40 μm and length $L = 19.5D_r$ independent of W. The channel is filled with fluid particles and with N suspended carriers and N_{RBC} RBCs. The number of RBCs is computed according to channel hematocrit, which corresponds to the area fraction of RBCs. The number of suspended particles for different simulations is provided in Table 1. An illustration of a typical simulation is shown in Fig. 2.

In the flow direction, periodic boundary conditions (BCs) are imposed, while in the other direction the suspension is confined by walls. The walls are modeled by frozen fluid particles with the same structure as the fluid, while the wall thickness is equal to r_c. Thus, the interactions of fluid particles with wall particles are the same as the interactions between fluid particles, and the interactions of suspended carriers and cells with the wall are identical to those with a suspending fluid. To prevent wall penetration, fluid particles as well as vertices of RBCs and carriers are subject to reflection at the fluid-solid interface. We employed bounce-back reflections, because they provide a better approximation for the no-slip boundary conditions in comparison to specular reflection of particles. To ensure that no-slip boundary conditions are strictly satisfied, we also add a tangential adaptive shear force [33] which acts on the fluid particles in a near-wall layer of a thickness $h_c = r_c$.

Blood flow is driven by a constant force applied to each solvent particle, which is equivalent to a prescribed pressure drop. To characterize the flow strength, we define a non-dimensional

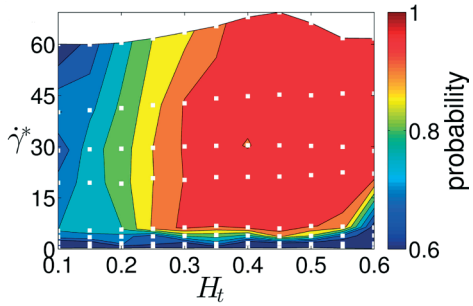


Fig. 4: *Probability diagram of particle margination with respect to $\dot{\gamma}^*$ and H_t, where the margination probability is defined as a probability of a particle center-of-mass to be within the RBCFL. The white squares (□) indicate the values of H_t and $\dot{\gamma}^*$ for which simulations have been performed.*

shear rate as

$$\dot{\gamma}^* = \bar{\dot{\gamma}}\tau = \bar{\dot{\gamma}}\frac{\eta D_r^3}{\kappa}, \tag{4}$$

where $\bar{\dot{\gamma}} = \bar{v}/W$ is the average shear rate (or pseudo shear rate) and $\bar{v}$ is the average flow velocity computed from the flow rate, while τ defines a characteristic RBC relaxation time.
To determine the RBC-free-layer (RBCFL) thickness, we measure the outer edge of the RBC core, which is similar to RBCFL measurements in experiments [34, 35]. The data is averaged for many RBC snapshots at different times.

3 Results

To study micro- and nano-particle margination for a wide range of conditions, we exploit the 2D blood flow model due to its numerical efficiency. However, our recent results [36] show that the 2D model is able to qualitatively reproduce the required blood flow characteristics and the particle margination effect in comparison with a realistic 3D model.

3.1 Particle margination

Carrier positions in blood flow sampled over time lead to particle distributions, which reflect the probability of a particle to be at a certain distance from the wall. Figure 3 shows several center-of-mass distributions of circular particles with $D_p = 0.3D_r$ (1.83 µm) for several H_t values and $\dot{\gamma}^* \approx 29.3$. The RBC-free layer (RBCFL) thickness is depicted by small arrows. The distributions have been averaged over the halves of the channel due to symmetry. Figure 3 shows that the carriers migrate into the RBCFL and remain quasi-trapped there. With increasing H_t, the carriers marginate better, as indicated by the development of a strong peak in the distribution near the wall at $y/W = 0$, and the motion of the peak position towards the wall. This is due to a decrease in the RBCFL thickness leading to a smaller available space for the particles. This trend is in agreement with experimental observations [14] and simulations [37, 38] of margination of blood platelets, which have a comparable size.

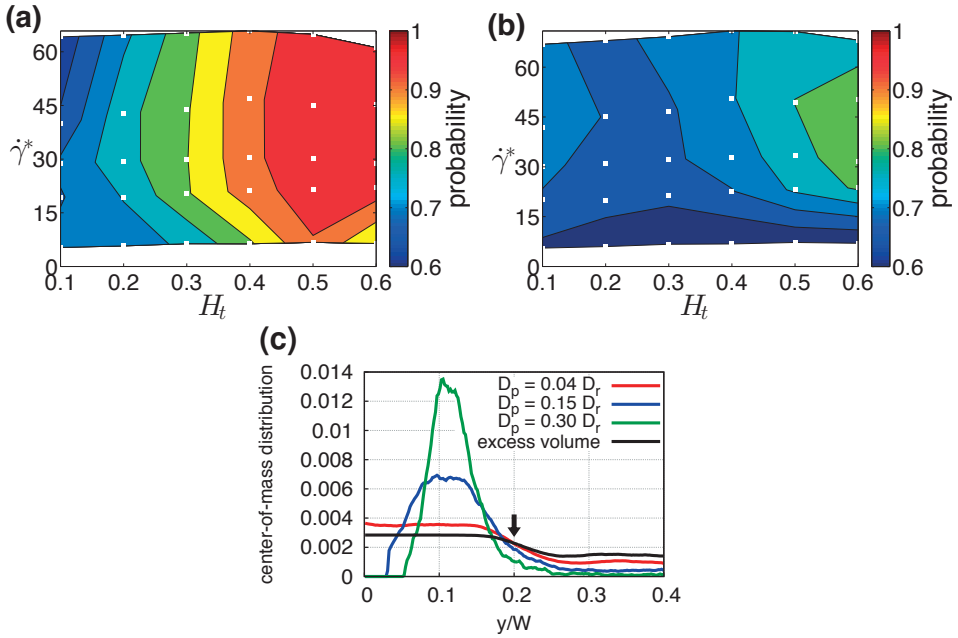


Fig. 5: *Dependence of margination on particle size. Probability diagrams of particle margination for various H_t and $\dot{\gamma}^*$ values and for circular particles with the sizes (a) $D_p = 0.15D_r$ (0.91 μm), (b) $D_p = 0.04D_r$ (0.25 μm). The white squares (□) indicate the values of H_t and $\dot{\gamma}^*$ for which simulation were performed. The margination probability is calculated based on the RBCFL thickness. (c) Distribution of particles with different sizes across the channel for $H_t = 0.3$ and $\dot{\gamma}^* \approx 29.3$. For small particles the distribution resembles the black solid curve computed as the blood-plasma volume. The arrow denotes position of the RBCFL boundary.*

To quantify and compare particle margination for a wide range of flow and particle parameters, we define the margination probability as a fraction of particles whose center-of-mass is located within the near-wall layer of thickness δ. The choice of δ depends on the exact problem to be addressed, and several possibilities can be considered. To describe particle margination into the vicinity of a vessel wall, it is natural to select δ to be the RBCFL thickness. Figure 4 presents margination probability diagram of particles for a wide range of H_t and $\dot{\gamma}^*$ values. Particle margination strongly depends on H_t as well as on shear rate. At low H_t values, particle margination is expected to be weak, while at high H_t the margination might be also attenuated due to particle-RBC interactions near a wall. The latter effect has been described for a marginating white blood cell [30] and is expected to subside for particles substantially smaller than a RBC, i.e. of sub-micrometer size. A pronounced dependence of particle margination on shear rate is observed at low flow rates. In the limit of very small flow rates ($\dot{\gamma}^* \lesssim 1$), the RBC distribution should be nearly uniform, and therefore, the RBCFL and consequently particle margination should almost vanish. As the shear rate is increased, the RBCFL thickness grows rapidly, leading to a substantial increase in particle margination.

The simulated values of $\dot{\gamma}^*$ cover the range of flow rates characteristic for the venular part of microcirculation ($\bar{\dot{\gamma}} \lesssim 80$ s^{-1} for $W \approx 20$ μm), where it is estimated that $\dot{\gamma}^* \lesssim 77$, while in arteriolar part the flow rates are higher ($\bar{\dot{\gamma}} \gtrsim 110$ s^{-1} for $W \approx 20$ μm) [39, 40]. The

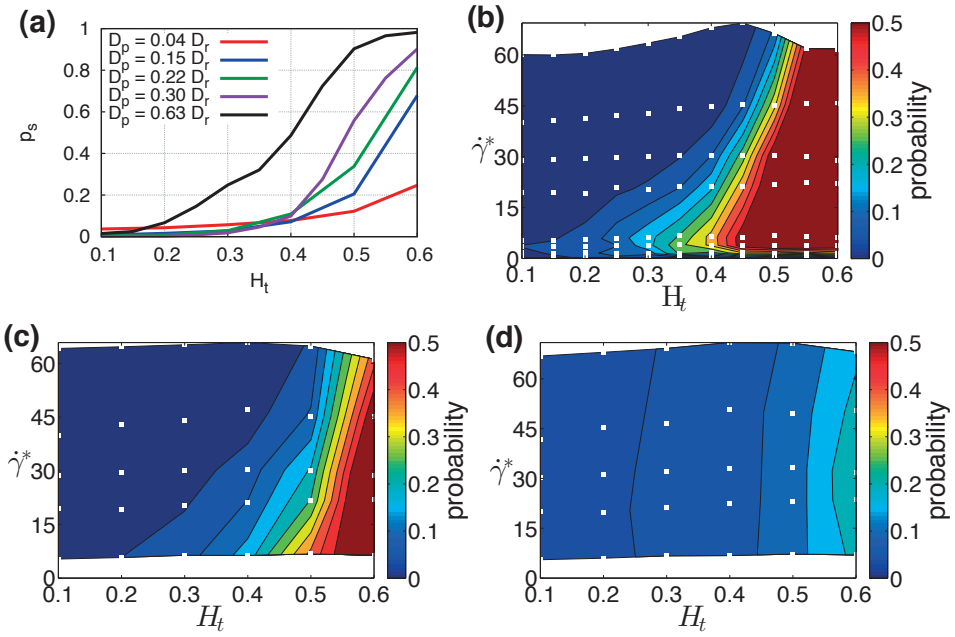


Fig. 6: *Margination into a potential adhesion layer of thickness* 200 *nm. (a) Margination probability* p_s*. The curves correspond to different particle sizes, where* $D_p = 0.63D_r$ *(*3.84 *μm) is for an elliptic particle and the other curves are for circular particles. Simulation results for* $\dot{\gamma}^* \approx 29.3$*. (b-d) Margination diagrams for (b)* $D_p = 0.3D_r$ *(*1.83 *μm) (c)* $D_p = 0.15D_r$ *(*0.91 *μm), and (d)* $D_p = 0.04D_r$ *(*0.25 *μm).*

considered range of shear rates is also relevant for tumor microvasculature, since blood flow velocities in tumors are much reduced in comparison to those under normal conditions, due to high geometric resistance and vessel permeability [41, 42]. Furthermore, the margination probability diagram in Fig. 4 shows that the strongest particle margination occurs in the range of $H_t = 0.2 - 0.6$. This region has a considerable overlap with the characteristic hematocrits in the body's microvascular networks in the range $H_t = 0.2 - 0.4$. A strong particle margination at high H_t values seems to be an advantage for drug delivery to tumors, since blood within tumor microvasculature is often subject to hemoconcentration due to plasma leakage [43].

3.2 Dependence of margination on particle size

The discussion above considered the margination of micron-size particles. There is also a strong interest in nano-carriers, with sizes starting from several nanometers. Figures 5(a),(b) show margination diagrams of particles with $D_p = 0.15D_r$ (0.91 μm) and $D_p = 0.04D_r$ (250 nm), respectively. The comparison of Figs. 5(a),(b) and Fig. 4 for $D_p = 0.3D_r$ (1.83 μm) reveals that the region of high margination probability becomes smaller with decreasing particle size. To illustrate the reason for the reduction in margination probability with decreasing particle size, we present in Fig. 5(c) the distributions of particles with different sizes for $H_t = 0.3$ and $\dot{\gamma}^* \approx 29.3$. For large enough particles, we observe a pronounced peak in the distribution next to the wall due to their interactions with RBCs, since their size is comparable with the RBCFL

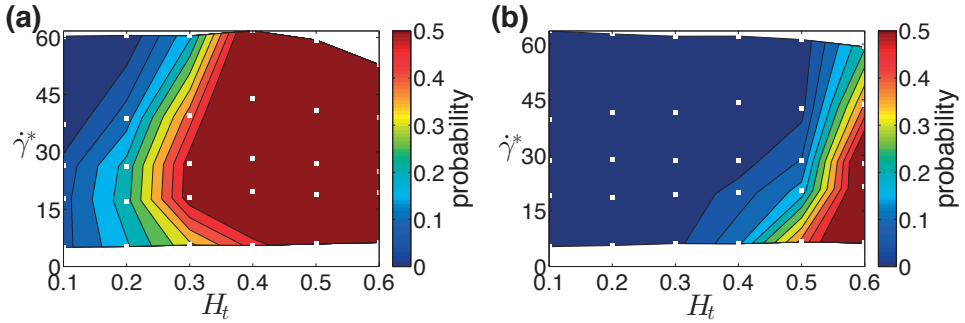


Fig. 7: *Margination for different channel widths. Margination into the potential adhesion layer based on* $\delta = 0.5D_p + 200$*nm, for particles with size* $D_p = 0.3D_r$ *(*$1.83\,\mu$*m) and two channel widths (a)* $W = 10\,\mu$*m and (b)* $W = 40\,\mu$*m.*

thickness. Even though small particles are also marginated, their distribution within the RBCFL is more uniform and their presence around the vessel center line is more probable than that for larger particles. Thus, the cumulative probability for a single particle to be within the RBCFL is lower for nano-carriers than that for micro-particles. Recent *in vivo* experiments [44] also support our numerical observations that particles with a size of about 1 µm are located closer to the vessel wall than smaller nano-particles. Noteworthy is that the distribution of the smallest particles with $D_p = 0.04D_r$ closely approaches the distribution computed as the excess fluid volume of flowing RBCs. This indicates that the distribution of particles smaller in size than roughly 250 nm can be well approximated by the distribution of the blood plasma, and therefore, their margination properties can be directly inferred from local H_t distributions.

To decide on a suitable particle size for efficient drug delivery, a number of different considerations have to be taken into account. A direct interpretation of probabilities in Figs. 4 and 5(a),(b) suggests that larger particle sizes are more favorable for drug delivery due to their better margination properties. To further support this proposition, we consider another definition for the margination probability based on $\delta = 0.5D_p + s$, which characterizes the fraction of carriers whose closest surface point is not further away from the wall than a distance s. We denote such a layer as "potential adhesion layer", since particle margination into a thin near-wall layer is a necessary precondition for adhesion. Even though the distance s is motivated by direct receptor-ligand interactions which occur within several nanometers, resolution restrictions in our mesoscale simulation approach do not allow the selection of smaller distances than approximately $s = 0.033D_r$, which corresponds to about 200 nm. Nevertheless, the distance of several hundred nanometers becomes relevant for particle-wall interactions in case of a carrier whose surface is decorated by tethered molecules [45]. Another definition for margination probability can also be based on a fixed layer thickness δ, thus it does not depend on H_t or on particle size. Figure 6(a) presents the margination probability into the potential adhesion layer (p_s) at $\dot{\gamma}^* \approx 29.3$. At very small H_t, the fraction of particles within the potential adhesion layer is small for all particle sizes; however, the smallest studied particles seem to be slightly more advantageous here. For the range of $H_t = 0.3 - 0.6$, Fig. 6(a) clearly shows that the fraction of large particles within the potential adhesion layer is much higher than that for small particles. The corresponding margination diagrams are shown in Figs. 6(b-d) and support the conclusion that large particles marginate better for all considered shear rates. This indicates that micro-carriers are likely to be better for drug delivery than sub-micron particles.

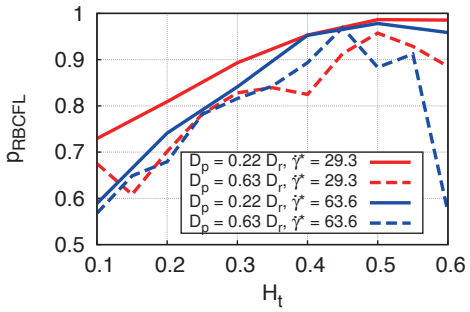


Fig. 8: *Dependence of margination on particle shape. Margination probabilities of ellipse-like particles (dashed lines) for various H_t and $\dot{\gamma}^*$ values in comparison to circular particles (solid lines) of the same area. The long axis of a 2D elliptic particle is $D_p = 0.63D_r$ (3.84 µm) and the aspect ratio equals approximately 7. The margination probability is calculated based on the RBCFL thickness.*

3.3 Dependence of margination on vessel size

To elucidate the effect of vessel diameter, we performed a number of simulations for two additional channel widths ($W = 10\,\mu$m and $40\,\mu$m) and two particle sizes ($D_p = 0.15D_r$ and $D_p = 0.3D_r$). The pronounced dependence of particle margination properties on channel width for the potential adhesion layer is illustrated by a comparison of Fig. 6(b) and Fig. 7. For particles with a size of $D_p = 0.3D_r$ (1.83 µm), particle margination into the potential adhesion layer improves considerably as the channel size decreases due to the much smaller RBCFL thickness in narrow channels. Thus, particle adhesion is expected to be more efficient in small vessels (i.e., capillaries) than in large vessels (i.e., venules and arterioles). Similar observation is found for particles with $D_p = 0.15D_r$ (0.91 µm) (not shown here). Furthermore, a reduction of margination into the potential adhesion layer with decreasing particle size is found for all channel sizes.

3.4 Dependence of margination on particle shape

Advances in micro- and nano-particle fabrication facilitate the production of carriers of various shapes, including spherical, prolate and oblate ellipsoidal, and rod-like shapes [4]. However, advantages of different particle shapes for drug delivery are still to be explored. Thus, we investigate the effect of shape on the margination properties in blood flow. Figure 8 displays results of simulations for the margination probability (based on the RBCFL) of elliptic particles under various blood flow conditions in comparison to circular particles. The ellipse has an aspect ratio of about 7 and the longest diameter is $D_p = 0.63D_r$ (3.84 µm); the enclosed area corresponds to the area of a circle with diameter $D_p = 0.22D_r$ (1.35 µm). The plot indicates that margination of elliptic particles is slightly worse than that of circular particles. From these data we can also conclude that margination of the elliptic particles with a smaller aspect ratio than 7 is similar to that presented in Fig. 8. However, since the largest diameter of the ellipse is larger than that of a circle with the same area, its margination into the potential adhesion layer, which is defined as a probability of a particle to be within a near-wall layer of thickness $\delta = 0.5D_p + 200$ nm, appears to be considerably larger for ellipses than that for the corresponding

circle (see Fig. 6). Recent theoretical [26, 27] and experimental [24, 25] studies also suggest that ellipsoidal particles possess better adhesion properties than spheres due to a larger contact area for adhesion interactions. In conclusion, the current knowledge about adhesion of ellipsoidal particles and our simulation results on margination suggest that ellipsoidal particles are very likely a better choice for drug delivery than spherical particles.

4 Discussion and conclusions

Particle margination in blood flow depends on particle size and shape, hematocrit, vessel size, and flow rate. Margination of circular and elliptical particles increases with increasing hematocrit, while their margination properties appear to be rather similar, where a circle marginates slightly more efficient than an ellipse. The presented diagrams show that larger particles have a higher margination probability in comparison to the smaller ones. Moreover, the distribution of very small particles with a diameter smaller than approximately 250 nm is well represented by the blood plasma volume of RBCs. Margination of particles into the potential adhesion layer is found to be more pronounced in small vessels, indicating that particle adhesion is likely to occur more often in capillaries than in arterioles and venules.

The simulation results are in good qualitative agreement with several experimental observations [13, 14, 21, 24, 25, 44]. For example, margination of micro-particles has been observed to be more efficient than that of nano-particles in recent *in vivo* experiments [44]. However, a detailed quantitative comparison is still difficult due to two reasons. On the one hand, the simulation results are obtained for 2D systems, which provide interesting insights into the relevant mechanisms, but have limited power for quantitative predictions for 3D systems. On the other hand, experimental data on particle margination in blood flow [13, 44] are very scarce and most of the available experimental investigations (e.g., Refs. [21, 24, 25]) focus on carrier adhesion. Even though margination is a necessary pre-condition for particle adhesion to vessel walls, particle margination and adhesion are not equivalent, since carrier adhesion may also depend on other factors (e.g., specific targets, the receptor/ligand density and distribution).

Clearly, the size and shape of drug carriers are important parameters not only for margination, but also for their adhesion and further transport through biological barriers (e.g., internalization). Our simulations suggest that elliptical particles are expected to adhere more efficiently than circular carriers due to a larger surface for adhesive interactions; however, this needs to be systematically investigated. Further requirements for efficient drug delivery include particle transport through vessel walls, interstitial space, and cell membranes. For instance, particle internalization by endothelial cells and intracellular trafficking have been shown to be most efficient for spherical sub-micron particles, rather than for micron-size carriers with an ellipsoidal shape [8]. This observation points in the direction of smaller carrier to be most efficient for internalization. As a consequence, the concept of multi-stage drug-delivery carriers [1, 5], where a larger micro-particle incorporates a number of small nano-carriers, seems to be very promising. In this way, margination and carrier delivery or adhesion to a specific target within the microvasculature could be achieved using micro-particles, which would then be followed by the release of nano-particles into the tissue. In conclusion, tackling various drug-delivery challenges is a complex issue; its resolution requires an inter-disciplinary effort including *in vitro* and *in vivo* experiments and realistic numerical simulations.

	m	a	γ	r_c	s	n	k_BT	η
$\dot{\gamma}^* \leq 20$	1	40	10	1.5	0.3	5	1	72.2
$\dot{\gamma}^* > 20$	1	40	20	1.5	0.3	5	1	144.4

Table 2: *DPD fluid parameters used in simulations. m is the mass of a fluid particle, a and γ are the conservative and dissipative force coefficients, respectively. r_c is the interaction cutoff radius, k is an exponent for the random-force weight function, n is the number density of fluid particles, k_BT is the energy unit with k_B being the Boltzmann constant and T temperature, and η is the fluid's dynamic viscosity. $\dot{\gamma}^*$ is the non-dimensional shear rate defined in the main text.*

Appendices

A Dissipative particle dynamics

Dissipative particle dynamics (DPD) [28, 29] is a mesoscopic particle-based method, where each particle represents a *molecular cluster* rather than an individual atom, and can be thought of as a soft lump of fluid. The DPD system consists of N point particles of mass m_i, position $\mathbf{r}_i$ and velocity $\mathbf{v}_i$. DPD particles interact through three forces: conservative ($\mathbf{F}^C_{ij}$), dissipative ($\mathbf{F}^D_{ij}$), and random ($\mathbf{F}^R_{ij}$) forces given by

$$\mathbf{F}^C_{ij} = F^C_{ij}(r_{ij})\hat{\mathbf{r}}_{ij}, \qquad \mathbf{F}^D_{ij} = -\gamma\omega^D(r_{ij})(\mathbf{v}_{ij}\cdot\hat{\mathbf{r}}_{ij})\hat{\mathbf{r}}_{ij}, \qquad \mathbf{F}^R_{ij} = \sigma\omega^R(r_{ij})\frac{\xi_{ij}}{\sqrt{dt}}\hat{\mathbf{r}}_{ij}, \tag{5}$$

where $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/r_{ij}$, and $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$. The coefficients γ and σ define the strength of dissipative and random forces, respectively. In addition, ω^D and ω^R are weight functions, and ξ_{ij} is a normally distributed random variable with zero mean, unit variance, and $\xi_{ij} = \xi_{ji}$. All forces are truncated beyond the cutoff radius r_c. The conservative force is given by

$$F^C_{ij}(r_{ij}) = a_{ij}(1 - r_{ij}/r_c) \quad \text{for} \quad r_{ij} \leq r_c, \tag{6}$$

where a_{ij} is the conservative force coefficient between particles i and j. The random and dissipative forces form a thermostat and must satisfy the fluctuation-dissipation theorem in order for the DPD system to maintain equilibrium temperature T [29]. This leads to

$$\omega^D(r_{ij}) = \left[\omega^R(r_{ij})\right]^2, \qquad \sigma^2 = 2\gamma k_B T, \tag{7}$$

where k_B is the Boltzmann constant. The choice for the weight functions is as follows

$$\omega^R(r_{ij}) = (1 - r_{ij}/r_c)^k \quad \text{for} \quad r_{ij} \leq r_c, \tag{8}$$

where k is an exponent. The time evolution of velocities and positions of particles is determined by the Newton's second law of motion

$$d\mathbf{r}_i = \mathbf{v}_i dt, \qquad d\mathbf{v}_i = \frac{1}{m_i}\sum_{j\neq i}\left(\mathbf{F}^C_{ij} + \mathbf{F}^D_{ij} + \mathbf{F}^R_{ij}\right)dt. \tag{9}$$

The above equations of motion are integrated using the velocity-Verlet algorithm [46]. The DPD fluid parameters used in simulations are given in Table 2.

References

[1] M. Ferrari, Nat. Rev. Cancer **5**, 161 (2005).

[2] R. K. Jain and T. Stylianopoulos, Nat. Rev. Clin. Oncol. **7**, 653 (2010).

[3] R. Duncan, Nat. Rev. Cancer **6**, 688 (2006).

[4] J. P. Rolland, B. W. Maynor, L. E. Euliss, A. E. Exner, G. M. Denison, and J. M. DeSimone, J. Am. Chem. Soc. **127**, 10096 (2005).

[5] E. Blanco, A. Hsiao, A. P. Mann, M. G. Landry, F. Meric-Bernstam, and M. Ferrari, Cancer Sci. **102**, 1247 (2011).

[6] W. R. Sanhai, J. H. Sakamoto, R. Canady, and M. Ferrari, Nat. Nanotechnol. **3**, 242 (2008).

[7] R. K. Jain, J. Natl. Cancer Inst. **81**, 570 (1989).

[8] S. Muro, C. Garnacho, J. A. Champion, J. Leferovich, C. Gajewski, E. H. Schuchman, S. Mitragotri, and V. R. Muzykantov, Mol. Ther. **16**, 1450 (2008).

[9] R. B. Huang, S. Mocherla, M. J. Heslinga, P. Charoenphol, and O. Eniola-Adefeso, Mol. Membr. Biol. **27**, 312 (2010).

[10] B.-S. Ding, T. Dziubla, V. V. Shuvaev, S. Muro, and V. R. Muzykantov, Mol. Interv. **6**, 98 (2006).

[11] U. Bagge and R. Karlsson, Microvasc. Res. **20**, 92 (1980).

[12] H. L. Goldsmith and S. Spain, Microvasc. Res. **27**, 204 (1984).

[13] G. J. Tangelder, H. C. Teirlinck, D. W. Slaaf, and R. S. Reneman, Am. J. Physiol. **248**, H318 (1985).

[14] A. W. Tilles and E. C. Eckstein, Microvasc. Res. **33**, 211 (1987).

[15] E. C. Eckstein, A. W. Tilles, and F. J. Millero III, Microvasc. Res. **36**, 31 (1988).

[16] I. Cantat and C. Misbah, Phys. Rev. Lett. **83**, 880 (1999).

[17] M. Abkarian, C. Lartigue, and A. Viallat, Phys. Rev. Lett. **88**, 068103 (2002).

[18] A. Kumar and M. D. Graham, Phys. Rev. Lett. **109**, 108102 (2012).

[19] J. D. Slack, M. Kanke, G. H. Simmons, and P. P. DeLuca, J. Pharm. Sci. **70**, 660 (1981).

[20] M. Herant, V. Heinrich, and M. Dembo, J. Cell Sci. **119**, 1903 (2006).

[21] K. Namdee, A. J. Thompson, P. Charoenphol, and O. Eniola-Adefeso, Langmuir **29**, 2530 (2013).

[22] D. C. Litzinger, A. M. J. Buiting, N. van Rooijen, and L. Huang, Biochim. Biophys. Acta **1190**, 99 (1994).

[23] W. H. De Jong, W. I. Hagens, P. Krystek, M. C. Burger, A. J. A. M. Sips, and R. E. Geertsma, Biomaterials **29**, 1912 (2008).

[24] F. Gentile, C. Chiappini, D. Fine, R. C. Bhavane, M. S. Peluccio, M. M.-C. Cheng, X. Liu, M. Ferrari, and P. Decuzzi, J. Biomech. **41**, 2312 (2008).

[25] A. J. Thompson, E. M. Mastria, and O. Eniola-Adefeso, Biomaterials **34**, 5863 (2013).

[26] P. Decuzzi and M. Ferrari, Biomaterials **27**, 5307 (2006).

[27] S. Dasgupta, T. Auth, and G. Gompper, Soft Matter **9**, 5473 (2013).

[28] P. J. Hoogerbrugge and J. M. V. A. Koelman, Europhys. Lett. **19**, 155 (1992).

[29] P. Español and P. Warren, Europhys. Lett. **30**, 191 (1995).

[30] D. A. Fedosov, J. Fornleitner, and G. Gompper, Phys. Rev. Lett. **108**, 028104 (2012).

[31] D. A. Fedosov, B. Caswell, and G. E. Karniadakis, Biophys. J. **98**, 2215 (2010).

[32] S. Suresh, J. Spatz, J. P. Mills, A. Micoulet, M. Dao, C. T. Lim, M. Beil, and T. Seufferlein, Acta Biomaterialia **1**, 15 (2005).

[33] D. A. Fedosov and G. E. Karniadakis, J. Comp. Phys. **228**, 1157 (2009).

[34] N. Maeda, Y. Suzuki, J. Tanaka, and N. Tateishi, Am. J. Physiol. **271**, H2454 (1996).

[35] S. Kim, R. L. Kong, A. S. Popel, M. Intaglietta, and P. C. Johnson, Am. J. Physiol. **293**, H1526 (2007).

[36] K. Müller, D. A. Fedosov, and G. Gompper, Sci. Rep. **4**, 4871 (2014).

[37] L. Crowl and A. L. Fogelson, J. Fluid Mech. **676**, 348 (2011).

[38] D. A. Reasor Jr, M. Mehrabadi, D. N. Ku, and C. K. Aidun, Ann. Biomed. Eng. **41**, 238 (2013).

[39] A. S. Popel and P. C. Johnson, Annu. Rev. Fluid Mech. **37**, 43 (2005).

[40] A. R. Pries, T. W. Secomb, and P. Gaehtgens, Am. J. Physiol. **269**, H1713 (1995).

[41] R. K. Jain, Cancer Res. **48**, 2641 (1988).

[42] F. Yuan, H. A. Salehi, Y. Boucher, U. S. Vasthare, R. F. Tuma, and R. K. Jain, Cancer Res. **54**, 4564 (1994).

[43] E. M. Sevick and R. K. Jain, Cancer Res. **49**, 3513 (1989).

[44] T.-R. Lee, M. Choi, A. M. Kopacz, S.-H. Yun, W. K. Liu, and P. Decuzzi, Sci. Rep. **3**, 2079 (2013).

[45] N. W. Moore and T. L. Kuhl, Biophys. J. **91**, 1675 (2006).

[46] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Clarendon Press, New York, 1991).

F 7 Microswimmers

Gerhard Gompper
Theoretical Soft Matter and Biophysics
Institute of Complex Systems, and Institute for Advanced Simulation
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**
1.1 Biological Microswimmers 2
1.2 Synthetic and Biomimetic Microswimmers 8
1.3 Model Microswimmers 10

2 Life at Low Reynolds Numbers **12**
2.1 Swimming at Low Reynolds Numbers 12
2.2 Microswimmer Flow Fields and Hydrodynamic Interactions 13

3 Swimming due to Flagellar Motion **14**
3.1 Anisotropic Hydrodynamic Friction of Slender Bodies 14
3.2 Swimming Velocity of Beating Flagella and Sperm 16

4 Active Brownian Particles **17**
4.1 Persistent Random Motion 18
4.2 Active Brownian Spheres near Surfaces 18
4.3 Active Brownian Rods near Surfaces 20
4.4 Collective Aggregation of Active Brownian Spheres 22

5 Hydrodynamic Synchronization **23**
5.1 Synchronization of Cilia in Metachronal Waves 24

6 Summary and Conclusions **26**

A Oseen Tensor of Low-Reynolds-Number Hydrodynamics **27**

B Hydrodynamic Force Dipoles **29**

C Hydrodynamics near Planar Surfaces **29**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Locomotion is an essential part of life. The search for food, for convenient environmental conditions (like temperature or acidity), and for partners for mating and reproduction, and the bisexual reproduction process itself all require locomotion. This is the case in the world of our daily experience, but also in the world of micro-organisms, such as cells, bacteria, and algae. Since these micro-organisms often live in an aqueous environment, locomotion means swimming.
Recently, there have been several attempts to construct artificial microswimmers. These swimmers sometimes mimic biological micro-organisms in their swimming mechanisms, sometimes employ new types of swimming motion. Synthetic microswimmers are interesting as future microrobots to perform tasks on the micrometer scale – such as directed motion in a body in medical applications. They are also interesting model systems to study the behavior of microswimmers without biological complications (such as an active response to external stimuli).

From the physics point of view, the field of microswimmers has been pioneered by Purcell [1], who showed that swimming in the microworld is quite different from our macroscopic experience, because at the length scale of nano- and micrometers, swimming occurs effectively at very high viscosity – for us like swimming in honey. During the more than 30 years, which have passed since Purcell published his article [1], the understanding of swimming microorganisms has increased extensively. This has lead to an exponential growth of publications in this field. Recent review articles provide a good overview of the present state of research and the current research directions. Generic aspects of the emergent large-scale behavior of self-propelled particles and active soft matter have been discussed in Refs. [2–7]. The hydrodynamics of swimming has been reviewed in Refs. [7–12]. Aspects of bacterial motility have been addressed in Refs. [7,13,14]. An overview of sperm motility and chemotaxis has been given in Ref. [15]. The dynamical properties of active Brownian particles have been discussed in Refs. [7, 16]. Finally, the propulsion of synthetic swimmer on the nanoscale and the development of nanomachines have been addressed in Refs. [17–19].

1.1 Biological Microswimmers

Bacteria. Some bacteria, like Escherichia coli (E. coli) [20] and Salmonella, swim by rotating helical hairlike filaments called flagella. Several of such flagella are attached to the bacterial body, as shown in Fig. 1. The flagella are rotated by a motor complex, which consists of several proteins, and which is anchored in the bacterial cell wall [20–22]; the motor is connected to the flagellum by a flexible hook, see Fig. 2.
Bacteria like E. coli and salmonella swim in a "run-and-tumble" motion illustrated in Fig. 3. In the "run" phase (stage 1 in Fig. 3), the helical winding of all flagella is left-handed, and they are rotating counter-clockwise. The flagella form a bundle (see also Fig. 1), and the bacterium moves forward in a direction determined by its long axis. At the beginning of the "tumble" phase, one flagellum reverses its rotational direction to clockwise (stage 2 in Fig. 3). This flagellum leaves the bundle, which implies a random reorientation of the bacterial orientation (stage 3 and 4). The reversal of the rotational direction is accompanied by a change of the helical handedness from left-handed to right-handed. At the end of the "tumbling" phase, all flagella start to rotate again in the same counter-clockwise direction (stage 5), the bundle reforms (stage 6), and the bacterium returns to a directional motion (stage 7 and 8).

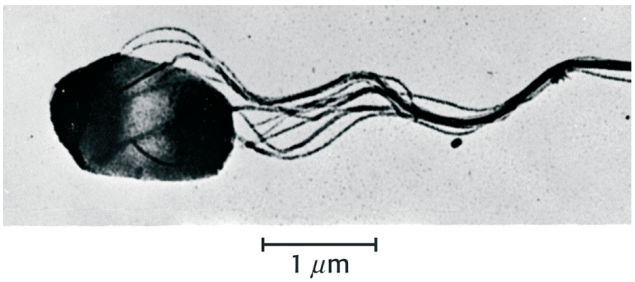


Fig. 1: *Salmonella swim by a bundle of rotating helical flagella. From Ref. [21].*

The purpose of the "run-and-tumble" motion is to detect gradients in the concentration of chemicals (e.g. food) or temperature (to avoid regions of too high or too low temperature). This is achieved by extending the "run" phase in case of improving environmental conditions, and by shortening it in case of worsening conditions.

Sperm. Sperm cells consist of a (roughly spherical) head, which contains the genetic material, a midpiece, which contains many mitochondria for energy production, and the eukaryotic flagellum, see Fig. 4. The structure of the flagellum is shown in Fig. 5. It consists of two central microtubules, which are surrounded by 9 double microtubules (microtubules are rather stiff filaments with a persistence length of about 1 mm). The microtubules are connected by many proteins (nexin links, central spokes, ...), which stabilize the structure. Motor proteins (dynein) connecting neighboring double microtubules cause an active bending of the structure by sliding the microtubules relative to each other.
A sperm cell is propelled though a fluid by a snake-like wiggling of its tail, as shown in Fig. 6. The flagellar beat is a propagating bending wave, with wave lengths smaller than the flagellar length. The beat can either be planar, as in Fig. 6, in particular for sperm swimming near surfaces [27], or be three-dimensional with a conical envelope.

Paramecium and Opalina. *Paramecia* is a group of unicellular ciliated protozoa, which range from about 50 μm to 350 μm in length. Many hair-like extrusions — called cilia — cover the body, see Fig. 7, which allow the cell to move with a synchronous motion (like a caterpillar) at speeds of approximately 2,700 μm/sec (12 body lengths per second). They generally feed on bacteria and other small cells. *Opalina* is a genus of protozoa found in the intestines of frogs and toads. It is without a mouth or contractile vacuole. The surface of Opalina is covered uniformly with cilia.
The structure of cilia and flagella of eukaryotic cells is very similar, see Fig. 5. The main structural difference between cilia and flagella is their length. A typical cilium is $10\mu m$ long, while sperm flagella are about $50\mu m$ long. The second large difference concerns their beat patterns. The ciliar beat has two distinct phases. During the *power stroke*, the cilium is stretched out straight and moves rather fast in one direction, while it bends, twists a little sideways and slowly retracts in the *recovery stroke*, as shown in Fig. 8. A particularly interesting feature of the beat pattern of cilia arrays on the surface of Opalina and Paramecium is the formation of "metachronal waves", as can be see in Fig. 7; the cilia do not beat synchronously, but in a pattern resembling a wheat field in the wind.
Different physical mechanisms for the beat of cilia and flagella have been suggested. Either the

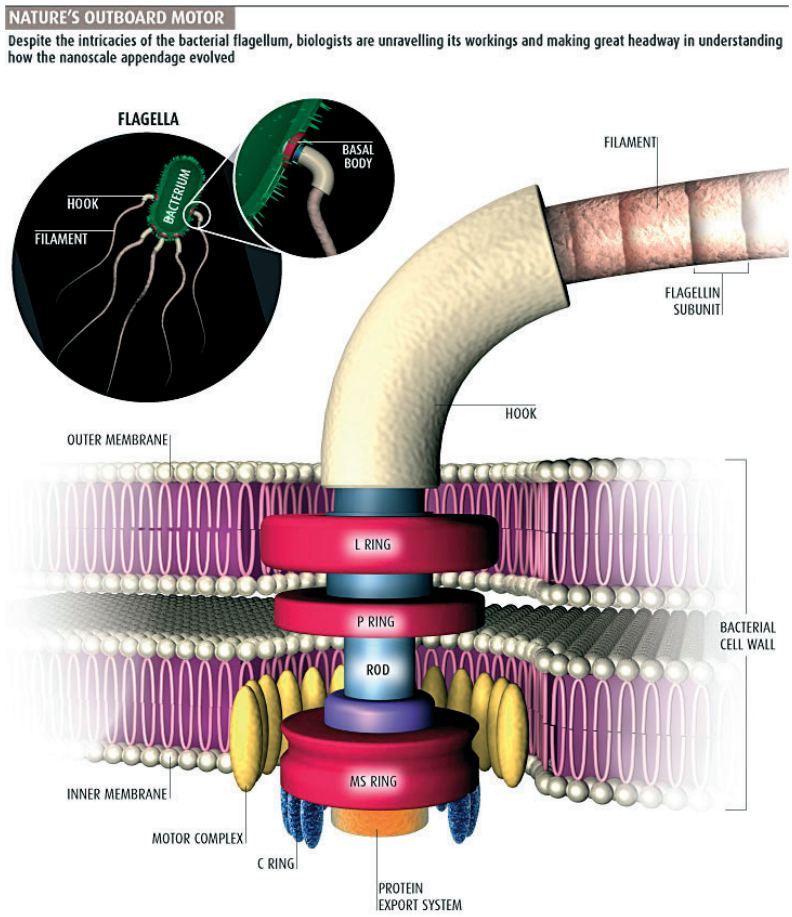


Fig. 2: *Each bacterial flagellum is driven by a rotary motor embedded in the bacterial cell wall. The motor has a series of rings, each about 20 nanometers in diameter, with a rod inside. Attached to the rod is a curved "hook" protein linked to the flagellum. The flagellum is 5 to 15 micrometers long and made of thousands of repeating units of the protein flagellin. The motor is powered by the flow of sodium or hydrogen ions across the cell wall. This generates rotation frequencies of up to 15 Hz. From Ref. [23].*

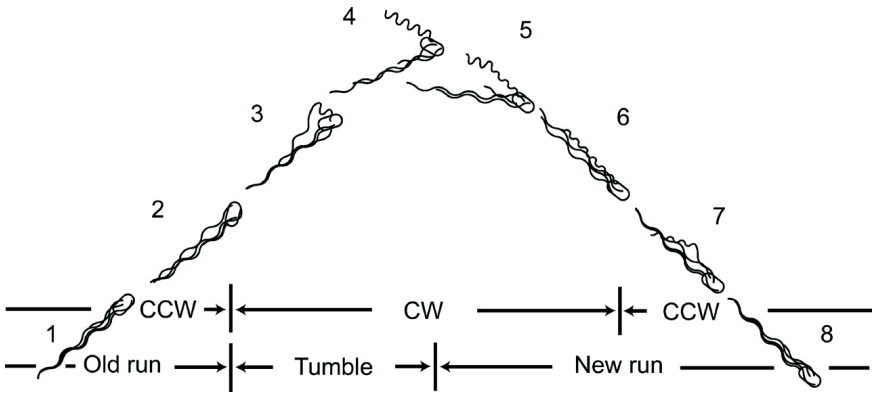


Fig. 3: *Bacteria like E. coli and salmonella move by a "run-and-tumble" motion. During the "run" phase, the flagella form a bundle, and the bacterium moves forward in one direction. In the "tumble" phase, one or a few flagella reverse their rotational direction and leave the bundle. This induces a tumbling motion which changes the orientation of the bacterium randomly. CW denotes clockwise, CCW counter-clockwise rotation. From Ref. [24].*

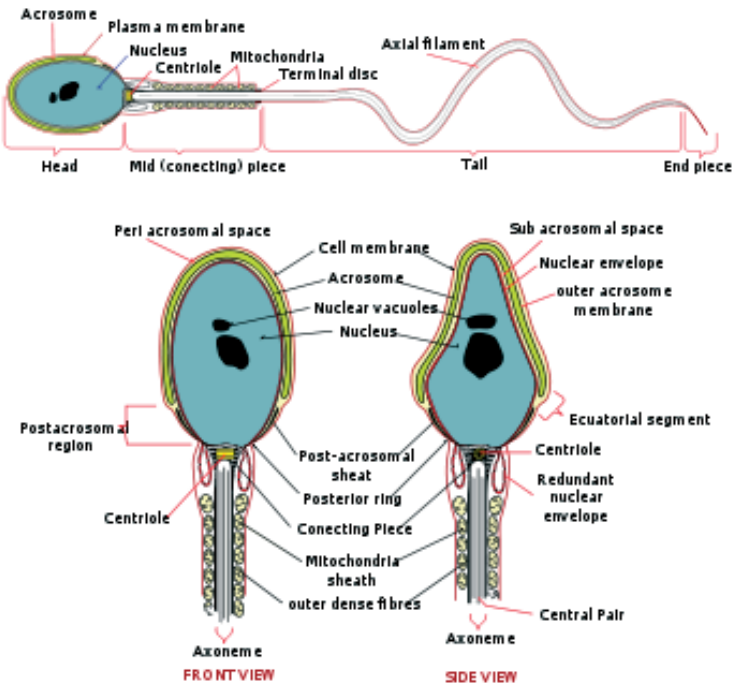


Fig. 4: *Schematic presentation of the structure of human sperm. Sperm cells consist of a head 5 μm by 3 μm and a tail 41 μm long. The tail flagellates, which propels the sperm cell at about 1–3 mm/minute by whipping in an elliptical cone. From Ref. [25].*

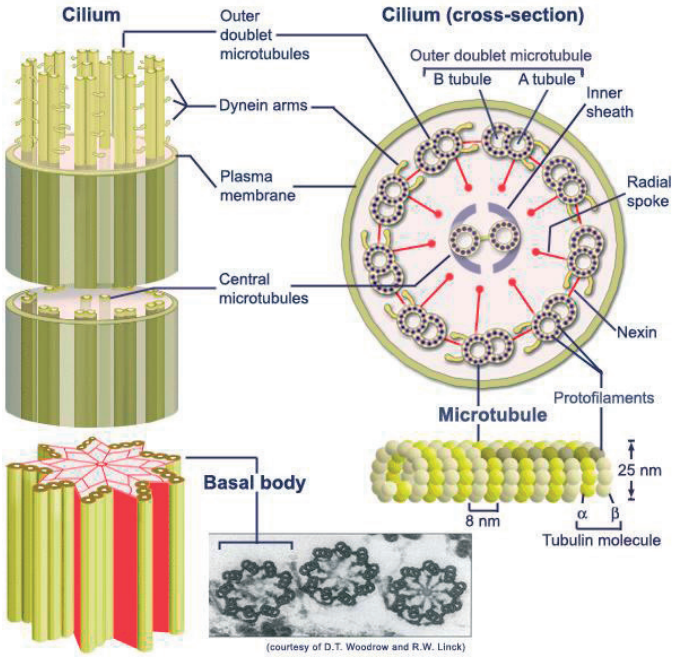


Fig. 5: *The axoneme consists of 9 double microtubules, arranged around the perimeter, and two central microtubules. The microtubules are connected by many linker proteins. Motor proteins connecting neighboring double microtubules lead to active bending. From Ref. [26].*

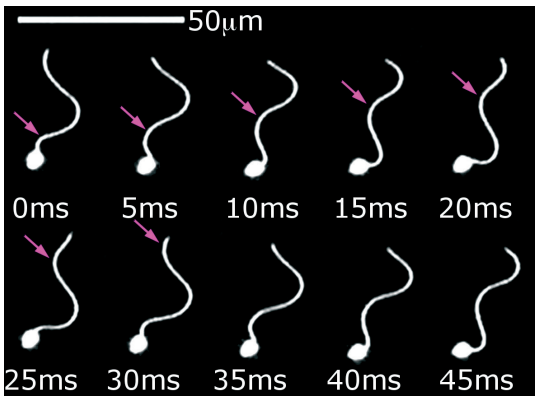


Fig. 6: *Sperm swim by a snake-like motion of their flagellum. The time sequence (from left to right) of snapshots of the beat of sea-urchin sperm shows a sinusoidal traveling wave on the flagellum. Arrow indicate the wave propagation from the head to the tip. From Ref. [7]*

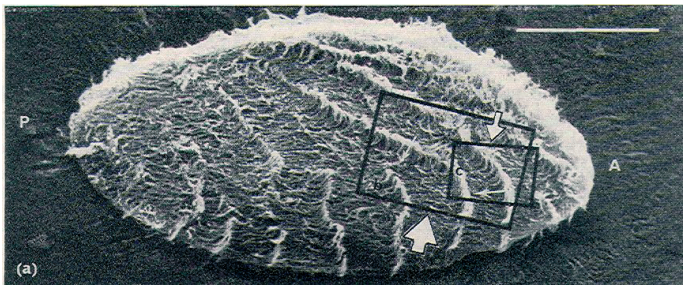


Fig. 7: *The surface of Opalina is covered by many hairlike filaments called cilia. Opalina swims forward by a stroke-like wiggling of the cilia. The formation of metachronal waves is clearly visible. From Ref. [28].*

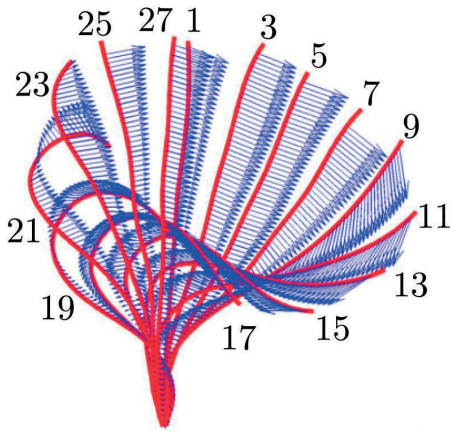


Fig. 8: *The beat of cilia has two distinct phases, the power stroke and the recovery stroke. Snapshots are shown from Volvox somatic cells, imaged at 1000 frames per second. During the power stroke, the cilium is stretched out straight and moves rather fast in one direction (frames 1 - 11), while during the recovery stroke, it bends and slowly retracts (frames 13-27). Adapted from Ref. [29].*

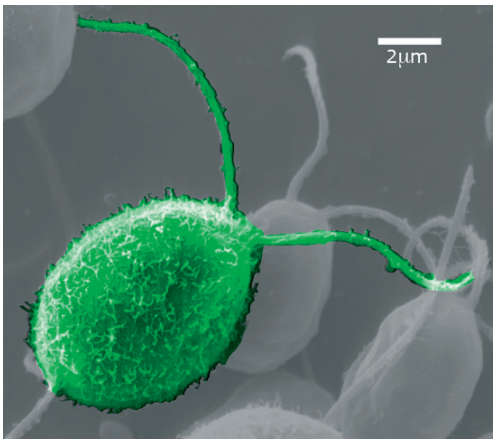


Fig. 9: Chlamydomonas *swims by a breast-stroke-like motion of its two flagella. Steering in reaction to a light stimulus is facilitated by additional beats of one flagellum. Adapted from Ref. [38].*

beat arises from a coupling of the activity of motor proteins with the curvature of the flagellum, where large curvature implies detachment of motors from the neighboring filament [30–32], or it arises from the cooperative behavior of several motors pulling in opposite directions, with the "winners" pulling along the "losers" for a while, as in a tug-of-war [33, 34].

Chlamydomonas reinhardtii. Chlamydomonas reinhardtii is a single-celled green algae, about 10 μm in diameter, that swims with two flagella, see Fig. 9. They have an "eye-spot" that senses light. When illuminated, C. reinhardtii can grow in a medium lacking carbon and energy sources. Widely distributed worldwide in soil and fresh water, C. reinhardtii is primarily used as a model organism in biology in a wide range of subfields [35]. The swimming motion of C. reinhardtii resembles the human breast-stroke: the flagella are pulled back in a nearly straight shape, and are then bend over and pushed forward again. The oscillatory velocity field induced by swimming C. reinhardtii in stabilized thin liquid films has been observed directly in time-resolved measurements recently [36, 37].

Volvox. Volvox is another green algae. It forms spherical colonies of up to 50,000 cells [39]. Each mature Volvox colony is composed of numerous flagellate cells similar to Chlamydomonas, embedded in the surface of a hollow sphere. The cells swim in a coordinated fashion, with distinct anterior and posterior poles. The cells have eye-spots, more developed near the anterior, which enable the colony to swim towards light. Recently, the flow field around freely swimming Volvox has been measured directly [36, 40].

1.2 Synthetic and Biomimetic Microswimmers

Locomotion on the nanoscale through a fluid environment is one of the grand challenges confronting nanoscience today [17]. The vision is to synthesize, probe, understand, and utilize a new class of motors made from nanoscale building blocks that derive on-board or off-board power from in-situ chemical reactions. The generated mechanical work allows these motors to

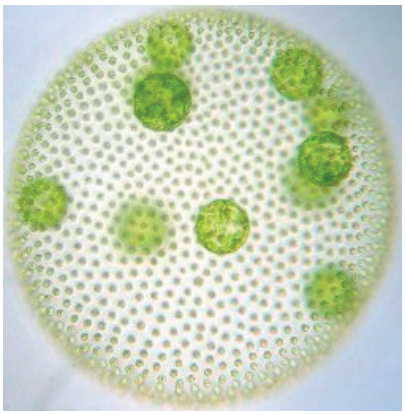


Fig. 10: *Volvox is a colony of several thousand of cells, which form a hollow sphere. Daughter colonies inside are visible. Each cell has two flagella, which beat in a coordinated fashion to rotate or move the colony. From Ref. [39].*

move through a fluid phase while simultaneously or sequentially performing a series of tasks. A large variety of such swimmers have been constructed recently, from bimetallic nanorods [17] and Janus colloids [41, 42] to artificial sperm [43, 44]. Some examples are given below.

Bimetallic Nanorods. A simple class of synthetic nanoscale motors are made from bimetallic Pt-Au nanorods immersed in a H_2O_2 solution [45, 46]. The catalytic reaction $2\,H_2O_2 \rightarrow 2\,H_2O + O_2$ occurs at the Pt end of the rod and is the power source for the motion. One plausible mechanism for motion involves the surface tension gradient due to O_2 adsorption on the nonreactive Au end. The molecular-level details of how O_2 generated at the Pt end of the nanorod leads to the propulsive force remain to be elucidated.

Diffusiophoretic Microspheres. Similarly, spherical non-conducting particles (like polystyrene or silica beads with metallic caps), which catalyze a chemical reaction inside the fluid, display self-propulsion [41, 42]. The catalytic reaction implies an asymmetric, non-equilibrium distribution of reaction products around the colloid, which generates osmotic or other phoretic forces [47–53]. These objects are denoted "diffusio-phoretic swimmers".

The concept of diffusio-phoretic swimmers can be taken one step further by constructing self-assembled, photoactivated colloidal microswimmers. An example is a polymer sphere, which includes a smaller hematite cube in dilute $\mathrm{H_2O_2}$ solution. The decomposition reaction is catalyzed by the hematite cubes, but only under illumination. Pairs of spheres and cubes then self-assemble into self-propelled microswimmers at a surface [54]. The dynamic assembly results from a competition between self-propulsion of particles and an attractive interaction induced respectively by osmotic and phoretic effects activated by light.

Thermophoretic Microspheres. Janus colloids with a metallic cap can also display self-propulsion due to self-thermophoresis. In this case, the cap is heated by a laser beam, which generates a temperature difference between the two sides of the Janus particle; the colloid then diffuses in this temperature gradient [55–58]. This approach has been extended to thermophoretic Janus colloids in binary fluid mixtures (with an upper miscibility gap) near the

demixing critical point [59,60]. In this case, heating by the laser beam leads to the formation of a droplet of one phase which adheres to the cap. This approach has the advantage that a small laser power suffices to induce self-propulsion.

Biomimetic Microswimmers. Artificial microswimmers can be constructed by using similar design principles as those found in biological systems. A by now classical example is a swimmer which mimics the propulsion mechanism of a sperm cell [43]. The flagellum is constructed from a chain of magnetic colloid particles and is attached to a red blood cell, which mimics the sperm head. This artificial swimmer is set into motion by an alternating magnetic field, which generates a sidewise oscillatory deflection of the flagellum. However, the swimming motion is not the same as for a real sperm cell; the wiggling motion is more a wagging than a traveling sine wave, and generates a swimming motion toward the tail end, opposite to the swimming direction of sperm.

A more recent example of a biohybrid swimmer, which mimics the motion of sperm consists of a polydimethylsiloxane filament with a short, rigid head and a long, slender tail on which cardiomyocytes (heart-muscle cells) are selectively cultured [44]. The cardiomyocytes contract periodically and deform the filament to propel the swimmer. This is a true microswimmer because it requires no external force fields.

Also, biomimetic systems have been designed to use artificial cilia for transport. For example, externally-actuated semi-flexible strings, like chains of magnetic beads [43, 61, 62], self-assembled microtubule bundles [63], ionic-polymer metal composite actuators [64], and even semi-flexible micro-pillars mounted an a basis of piston-like actuators [65] have been proposed to employ the cilia propulsion mechanism in artificial nano-machines and microfluidic devices [66–68].

1.3 Model Microswimmers

Three-Bead Swimmer. It was recognized already more than 30 years ago by Purcell [1] that directed forward swimming of micromachines in viscous fluids — at low Reynolds numbers, see Sec. 2 below — is not possible when the angle between *two* rigid segments is varied periodically in time ("scallop theorem"). A time-irreversible internal motion is required, which needs at least two degrees of freedom. This can be achieved with a very simple one-dimensional swimmer, which is constructed from three spheres that are linked by rigid rods, whose lengths can change between two values [69]. With a periodic motion that breaks the time-reversal symmetry as well as the translational symmetry, the model device swims at low Reynolds number as shown in Fig. 11. This swimmer has the advantage that many of its properties can be analyzed in detail or even calculated analytically [69–71].

Squirmers. One class of microswimmers are almost spherical organisms that are propelled by active hair-like organelles (cilia) covering the body. On a mesoscopic length scale, the synchronized beating of the cilia can be mapped onto a spherical envelope [72,73], and its time average corresponds to a steady tangential surface velocity. These objects — called "squirmers" — may also serve as a simple generic model for other types of microswimmers, for example diffusio-phoretic Janus spheres [41,42].

The squirmer is modeled as a hard sphere of radius R with a prescribed tangential surface velocity $\mathbf{v}_{\rm sq}$, causing a propulsion in the direction of the squirmer's instantaneous orientation $\hat{\mathbf{e}}$.

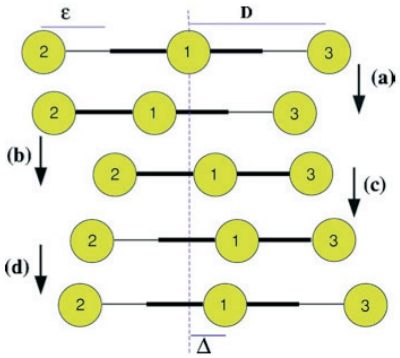


Fig. 11: *Three connected beads swim forward when their distances are varied cyclically as a function of time. Here, it is important that the cycle breaks time-reversal symmetry, as shown. In one cycle, the swimmer moves by a distance* Δ*. From Ref. [69].*

The relative velocity at a surface point $\mathbf{r}_\mathrm{s}$ (with respect to the squirmer's center) is given by [73]

$$\mathbf{v}_\mathrm{sq}(\mathbf{r}_\mathrm{s},\hat{\mathbf{e}}) = \sum_{n=1}^{\infty} B_n \frac{2}{n(n+1)} \left(\frac{\hat{\mathbf{e}}\cdot\mathbf{r}_\mathrm{s}}{R}\frac{\mathbf{r}_\mathrm{s}}{R} - \hat{\mathbf{e}} \right) P_n'\left(\frac{\hat{\mathbf{e}}\cdot\mathbf{r}_\mathrm{s}}{R} \right), \tag{1}$$

where $P_n'(\cos\theta)$ is the derivative of the *n*th Legendre polynomial and B_n is the amplitude of the *n*th mode of the surface velocity. The absolute local surface velocity of the squirmer is given by

$$\mathbf{v}_\mathrm{s}(\mathbf{r}-\mathbf{r}_\mathrm{c},\hat{\mathbf{e}}) = \mathbf{v}_\mathrm{c} + \mathbf{v}_\mathrm{sq}(\mathbf{r}-\mathbf{r}_\mathrm{c},\hat{\mathbf{e}}) + \boldsymbol{\Omega}\times(\mathbf{r}-\mathbf{r}_\mathrm{c}), \tag{2}$$

where $\mathbf{r}_\mathrm{c}$, $\mathbf{v}_\mathrm{c}$ and $\boldsymbol{\Omega}$ are the sphere position, velocity and angular velocity, respectively. The constant B_1 sets the average velocity of the squirmer, $v = \langle \hat{\mathbf{e}}\cdot\mathbf{v}_\mathrm{c}\rangle = \frac{2}{3}B_1$ [74]. To cope with the biodiversity of real microswimmers, the characteristic surface velocity of the model can be varied by adjusting the coefficients B_n with $n \geq 2$. In most cases studied so far, only B_1 and B_2 are taken to be non-zero, so that $B_n = 0$ for $n \geq 3$.
This model was employed, for example, to study the hydrodynamic interaction of two squirmers without [75] and with [76] thermal fluctuations, monolayers of squirmers without thermal fluctuations [77], and squirmers at high density in quasi-two-dimensional confinement [78].

Rotators. Rotating discs [79,80], rotating spheres or dumbbells [81] are another important class of swimmers. These rotators can be autonomous swimmers, like the Volvox algae mentioned above, or actuated synthetic particles, like super-paramagnetic particles rotated by an external magnetic field [82]. Rotators show an interesting collective behavior, like a circling motion of Volvox around each other [40], or a lane formation of rotating magnetic discs in microfluidic channels [83, 84].

Vicsek Model for Swarming. Collective behavior of active bodies is frequently found in microscopic systems such as bacteria [85–89] and synthetic microswimmers [54, 90–93], but also in macroscopic systems such as flocks of birds and schools of fish [4, 94]. Despite the very different propulsion mechanisms and interactions in these systems, they all favor alignment of neighboring bodies, thus leading to similar forms of collective behavior. Therefore, it is

natural to look for a model, which is able to capture the generic collective properties of all the various systems of self-propelled particles and organisms. Such a model was proposed in a pioneering work by Vicsek [95]. In this model, now often called the "Vicsek model", N polar point particles move in space with constant magnitude of velocity v_0. The dynamics proceeds in two steps, a streaming step of duration Δt, in which particles move ballistically, and a interaction step, in which particles align their velocity direction with the average direction of motion of their neighbors. In two spatial dimensions, this implies the dynamics for the position $\boldsymbol{r}_i$ and velocity $\boldsymbol{v}_i$ of particle i,

$$\boldsymbol{r}_i(t+\Delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t)\Delta t, \tag{3}$$

$$\theta_i(t+\Delta t) = \langle\theta(t)\rangle_\sigma + \Delta\theta, \tag{4}$$

where θ_i is the angle between $\boldsymbol{v}_i$ and the x-axis of a Cartesian coordinate system, $\langle\theta(t)\rangle_\sigma$ is the average orientation of all particles within a circle of diameter σ. and $\Delta\theta$ is a random noise uniformly distributed in the interval $[-\eta/2, +\eta/2]$. The essential control parameters of the Vicsek model are the particle density ρ, the noise strength η, and the propulsion velocity v_0 in units of $\sigma/\Delta t$. Numerical investigations of this model show a phase transition with increasing density or decreasing noise strength from a random isotropic phase to an aligned phase, in which particles move collectively in a spontaneously selected direction.

2 Life at Low Reynolds Numbers

2.1 Swimming at Low Reynolds Numbers

The dynamics of fluids — hydrodynamics — is described by the (incompressible) Navier-Stokes equation,

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right) = \eta\nabla^2\mathbf{v} - \nabla p + \mathbf{f}_{ext} \tag{5}$$

where ρ is the fluid density, η the fluid viscosity, $\mathbf{v}(\mathbf{r},t)$ the position- and time-dependent fluid velocity field, $p(\mathbf{r},t)$ the pressure field, and $\mathbf{f}_{ext}(\mathbf{r},t)$ an external body force. The Navier Stokes equation can be written in dimensionless form by scaling all lengths with a characteristic length L, velocities by a characteristic velocity v_0, and time by L/v_0. This implies

$$\mathfrak{Re}\left(\frac{\partial \mathbf{v}'}{\partial t'} + (\mathbf{v}'\cdot\nabla)\mathbf{v}'\right) = \nabla^2\mathbf{v}' - \nabla p' + \mathbf{f}'_{ext} \tag{6}$$

where the prime denotes dimensionless quantities. Here, $\mathfrak{Re}$ is the Reynolds number, which is found to be

$$\mathfrak{Re} = \rho v_0 L/\eta\ . \tag{7}$$

For microswimmers, the Reynolds number is typically very small, because the characteristic lengths and velocities are small. For a swimmer of length $L = 50\mu$m and a velocity of 10 body lengths per second (which is already a large velocity), the Reynolds number in water (with a kinematic viscosity $\eta/\rho = 10^{-6}m^2/s$) is $\mathfrak{Re} = 0.025$. In this case, the nonlinear contributions on the left-hand-side of Eq. (6) can be neglected, leading to the simpler Stokes equation,

$$\nabla p - \eta\nabla^2\mathbf{v} = \mathbf{f}_{ext} \tag{8}$$

Although this equation is much simpler than the Navier-Stokes equation, it is not possible to solve analytically for microswimmer motion – even in simple geometries.
At surfaces, the fluid velocity is typically very small, because the collisions of fluid molecules with the surface imply that the molecules are scattered backwards and thereby transfer momentum parallel to the wall. Thus, no-slip boundary conditions, with $\mathbf{v}(\mathbf{r}) = 0$ at the surface, are usually employed.
It has been recognized by Purcell [1] that life at low Reynolds numbers is quite different from the life of our everyday experience. Since low Reynolds numbers can either be achieved by small length scales (and small velocities) or by very high viscosities, a swimming microorganism experiences similar dynamics as a human swimming in honey. Since inertia effects are negligible, forward motion stops immediately when the "motor" stops working. Another important consequence of low Reynolds number is that the Stokes equation is time-reversible. Therefore, a time-reversible motion of a micro-organism will generate an oscillatory, but no directed motion. This has been called the "scallop theorem" by Purcell [1]. It means that just by opening and closing its two shells, a mussel cannot move forward. As explained above for the Purcell-swimmer, a second degree of freedom is required to generate a sequence of moves which is not time reversible.

2.2 Microswimmer Flow Fields and Hydrodynamic Interactions

Most microswimmers move autonomously, with no external force applied, and hence the total interaction force of the swimmer on the fluid, and *vice versa*, vanishes. In the simplest case, which actually applies to many microswimmers like bacteria, spermatozoa, or algae, the far-field hydrodynamics (at distances from the swimmer much larger than its size) can well be described by a force dipole [8, 9]. This has been confirmed experimentally for *E. coli* [96]. Two classes of such dipole swimmers can be distinguished, as shown schematically in Fig. 12. If the swimmer has its motor in the back, and the passive body drags along the surrounding fluid in front, the characteristic flow field of a "pusher" emerges, see Fig. 12(a). Similarly, if the swimmer has its motor in the front, and the passive body drags along the surrounding fluid behind, the characteristic flow field of a "puller" develops, see Fig. 12(b). Thus, sperm, E. coli and salmonella are pushers, because they have active propelling flagella in the rear, and a passive head in the front. In contrast, C. reinhardtii is a puller, because the flagella in the front push the fluid backwards, while the body in the rear remains passive. It is important to notice that the flow fields of pushers and pullers look similar, but with opposite flow directions. This has important consequences for the interactions between swimmers, and of swimmers with walls, as will be explained below.
A first idea about the effect of hydrodynamic interactions on the dynamics of microswimmers near surfaces can be obtained again from a far-field approximation, which applies when the size of the swimmer is much smaller than its distance from the surface, so that it can be approximated by a force dipole, compare Appendix C. In this limit, analytical expressions for the torque on a swimmer and its drag velocity towards the surface can be obtained [97]. For a swimmer with an orientation angle θ of the dipole with respect to the vertical (z) direction (i.e. $\theta = 90°$ when the swimmer moves parallel to the wall), the induced velocity at a distance z away from the no-slip wall is given by [97]

$$u_z(\theta, z) = -\frac{3P}{64\pi\eta z^2}\left(1 - 3\cos^2\theta\right). \tag{9}$$

As shown in Appendix B, the dipole strength P is given by the product of dipole length and

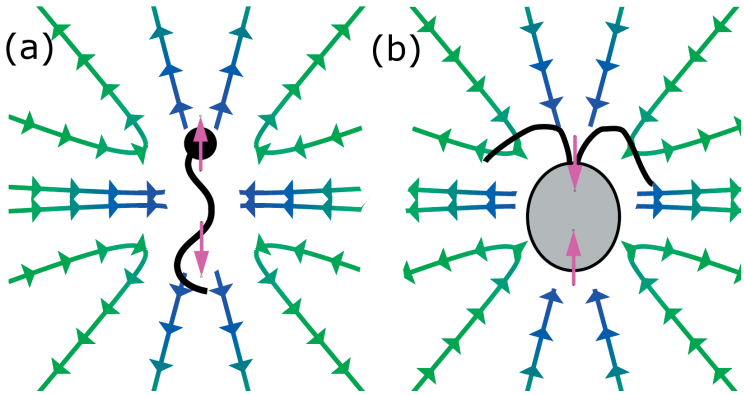


Fig. 12: *Schematics of the flow field of dipole swimmers, (a) pusher and (b) puller. From Ref. [7].*

force. The dipole force of sperm can be estimated from the friction force of the head, which — for a sufficiently large head — balances the pushing force of the tail, and is thus proportional to $\eta v L$ (compare Eqs. (14) and (17)), where v is the swimming velocity and L the sperm length. The dipole strength therefore is

$$P \sim \eta v L^2 \, . \tag{10}$$

Equation (9) allows several interesting predictions. First, the hydrodynamic interaction is attractive for swimmers oriented nearly parallel to the wall (with θ near $90°$), but becomes repulsive for swimmers oriented nearly perpendicular to the wall (with θ near $0°$). Second, the hydrodynamic interaction is long-ranged, with a $1/z^2$ decay with increasing distance from the surface. Finally, the hydrodynamic interaction has been shown not only to generate a force on the swimmer, but also a torque [97]. It turns out that this torque always acts to align the swimming cell parallel to the surface – the orientation in which the hydrodynamic force is attractive.
Similarly, the interaction of two microswimmers at long distances is determined by their dipole flow fields. The dipole approximation predicts that the interactions of pushers and pullers have opposite sign, because their dipole strengths P have opposite signs (compare Eq. (56) of Appendix C). This behavior can be seen explicitly by considering the flow fields of two parallel-swimming pushers or pullers [76]. The results of mesoscale simulations of two squirmers shown in Fig. 13 demonstrate that for *pushers*, the fast backward flow velocity in the rear part extracts fluid from the gap between the swimmers, and thereby induces attraction (Fig. 13a); in contrast, for *pullers*, the fast backward flow velocity in the front part injects fluid into the gap, and thereby induces repulsion (Fig. 13b). At the small squirmer separation show in Fig. 13, the interaction is dominated by the hydrodynamic near-field, and the far-field approximation does not allow quantitative predictions anymore.

3 Swimming due to Flagellar Motion

3.1 Anisotropic Hydrodynamic Friction of Slender Bodies

A microorganism is able to swim forward in a fluid by wiggling or rotating a flagellum, because the hydrodynamic friction of a long, slender body in a viscous environment is *anisotropic*. This

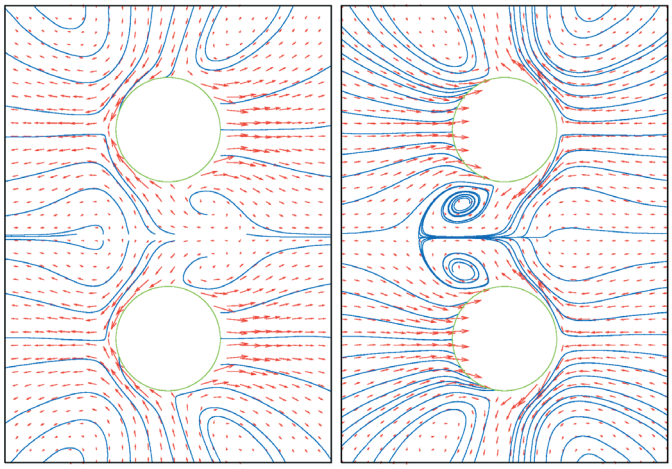


Fig. 13: *Velocity fields for fixed parallel pairs of squirmers, for (a) pusher (with $B_2/B_1 = -3$) and (b) puller (with $B_2/B_1 = +3$), with Péclet number $\mathrm{Pe} = 1155$. Swimmers move to the right. Streamlines serve as a guide to the eye. Only a fraction of the simulation box is shown. (Due to the finite resolution of the measured velocity field, some streamlines end on the squirmers' surfaces.) From Ref. [76].*

can be demonstrated easily for a long and thin rod of diameter d and length L: it experiences less friction when pulled along its axis than perpendicular to it.
Let us consider the motion of the rod, which may be oriented in the x-direction, with an applied force in direction $\hat{\mathbf{e}}$. We approximate the rod as a sequence of touching spheres of diameter d. The friction coefficient ζ of a sphere (with no-slip boundary conditions) moving in a viscous fluid under a force F is given by

$$\mathbf{F} = \zeta_s \mathbf{v} \ , \quad \text{where} \quad \zeta_s = 3\pi\eta d \tag{11}$$

is the Stokes friction coefficient. The dynamics of bead n in a chain with other beads is determined in general by the Langevin equation [98]

$$\mathbf{v}_n = \frac{\partial}{\partial t}\mathbf{R}_n = \sum_m \mathbf{H}_{nm} \cdot \left[-\frac{\partial U}{\partial \mathbf{R}_m} + f_m \right] \tag{12}$$

where $U(\mathbf{R}_1, \mathbf{R}_2, ...)$ is the interaction potential among the beads (including, in particular, the bonds to nearest neighbors), $\mathbf{H}_{nm}$ for $n \neq m$ is the Oseen tensor (see Eqs. (44) and (50) derived in Appendix A), $\mathbf{H}_{nn} = \mathbf{I}/\zeta_s$, and f_m is an external force on bead m. Lets us now assume that the force $\mathbf{f}$ on all beads is the same, and that the interaction between beads can be neglected (this is only valid for short times, but should not affect the calculation of the friction coefficient) [99]. Then,

$$\mathbf{v}_{rod} \equiv \mathbf{v}_n = \left[\mathbf{I}/\zeta_s + \sum_{m \neq n} \mathbf{H}_{nm} \right] \cdot \mathbf{f} \ . \tag{13}$$

The anisotropic friction coefficients of a rod are defined by

$$\mathbf{F} = \zeta_\parallel \mathbf{v}_\parallel + \zeta_\perp \mathbf{v}_\perp \ . \tag{14}$$

Calculations based in Eq. (14), with constant friction coefficients $\zeta_\parallel$ and $\zeta_\perp$, are denoted "resistive-force theory". To calculate $\zeta_\parallel$ and $\zeta_\perp$ for a rod, it is easiest to consider the special cases of a rod pulled parallel and perpendicular to its long axis. Then, with the sum replaced by an integral, for a rod oriented in the x-direction, Eq. (51) gives for the central bead

$$\mathbf{v}_{rod} \equiv \mathbf{v}_n = f \left[\frac{\hat{\mathbf{e}}}{\zeta_s} + \frac{2}{8\pi\eta d} \int_d^{L/2} \mathrm{dx} \, \frac{1}{x} \left(\hat{\mathbf{e}} + (\hat{\mathbf{e}}_x \cdot \hat{\mathbf{e}}) \hat{\mathbf{e}}_x \right) \right] . \tag{15}$$

with $\mathbf{f} = f\hat{\mathbf{e}}$. Because $(\hat{\mathbf{e}}_x \cdot \hat{\mathbf{e}})\hat{\mathbf{e}}_x = 1$ and 0 for parallel and perpendicular orientation, respectively, we find immediately that

$$\zeta_\perp = 2\zeta_\parallel \tag{16}$$

in the limit of very long rods. It is therefore a factor 2 easier to pull a long rod along its axis than perpendicular to it. Finally, the integration in Eq. (15) together with $F = (L/d)f$ gives

$$\zeta_\perp = \frac{4\pi\eta L}{\ln(L/d)} \tag{17}$$

to leading order in L/d. The logarithmic divergence is a result of the long-ranged hydrodynamic interaction of different parts of the rod, which *reduce* the friction coefficient compared to that of a chain of non-interacting beads.

3.2 Swimming Velocity of Beating Flagella and Sperm

The result (16) together with Eq. (14) can now be used to calculate the swimming velocity of an sinusoidally beating flagellum. In this case, the shape as a function of time t is described by

$$y(x,t) = A \sin(kx - \omega t) \tag{18}$$

where A is the beat amplitude, ω the beat frequency, and $k = 2\pi/\lambda$ the wave number. The velocity of a segment of the flagellum at x is then

$$v_y(x,t) = \frac{\partial y}{\partial t} = -A\omega \cos(kx - \omega t) \tag{19}$$

With the local tangent vector (not normalized)

$$\mathbf{t}(x,t) \sim (1, \partial y/\partial x, 0) = (1, Ak\cos(kx - \omega t), 0) \tag{20}$$

the velocity $\mathbf{v}(x,t) = (0, v_y(x,t), 0)$ can be decomposed into $\mathbf{v}_\parallel = (\mathbf{v}\cdot\mathbf{t})\mathbf{t}/\mathbf{t}^2$ and $\mathbf{v}_\perp = \mathbf{v} - \mathbf{v}_\parallel$ with

$$\mathbf{v}_\parallel = -\frac{A^2\omega k \cos(kx - \omega t)^2}{1 + A^2k^2\cos(kx - \omega t)^2} \, \mathbf{t} \tag{21}$$

According to Eq. (14), this generates a force F_x in the swimming direction with

$$F_x = (\zeta_\parallel - \zeta_\perp) \int \mathrm{dx} \frac{A^2\omega k \cos(kx - \omega t)^2}{1 + A^2k^2\cos(kx - \omega t)^2} \tag{22}$$

while the force in the perpendicular direction vanishes when averaged over the whole flagellum. For small beat amplitudes, Eq. (22) can easily be integrated to give

$$F_x = \frac{1}{2}(\zeta_\parallel - \zeta_\perp) A^2 \omega k \tag{23}$$

The swimming velocity then follows from $v_x \simeq F_x/\zeta_\parallel$ to be

$$v_{flag} = -\frac{1}{2}\left(\frac{\zeta_\perp}{\zeta_\parallel} - 1\right) A^2 \omega k \; . \tag{24}$$

This result shows several interesting features. First, it shows that swimming is only possible due to the friction anisotropy, since the velocity is proportional to $(\zeta_\parallel - \zeta_\perp)$. Second, for a traveling wave in the positive x-direction, the flagellum moves in the negative x-direction, i.e. movement is opposite to the direction of the traveling wave. Third, the swimming velocity increases linearly with the beat frequency ω and the wave vector k, but quadratically with the beat amplitude A. And finally, the swimming velocity is independent of the fluid viscosity.
A more refined calculation has been performed in Ref. [100], also employing resistive force theory, to determine the swimming velocity of sperm. For the same sinusoidal beat pattern as in Eq. (18) and $\zeta_\perp/\zeta_\parallel = 2$, this leads to [100]

$$v_{sperm} = \frac{1}{2}A^2\omega k \left[1 + A^2k^2 + \sqrt{1 + \frac{1}{2}A^2k^2}\frac{3r_h}{L_{flag}}(\ln(kd/4\pi) - 1/2)\right]^{-1} \tag{25}$$

Here L_{flag} is the length of the flagellum, and r_h is the radius of the head. The general conclusions of Eq. (24) remain valid, but additional effects appear. The second term in the brackets of Eq. (25) — whose origin is already recognizable in Eq. (21) — arises from the finite beat amplitude, and implies that the velocity *saturates* for large beat amplitudes A. The last term in the brackets describes the reduction of velocity due the drag of the passive head. The swimming of sperm has also been analyzed by slender-body theory (taking into account the hydrodynamic interactions of different parts of the deformed flagellum, as in Sec.3.1 for slender rods) [101]. Results agree with the resistive-force results of Ref. [100] within about 10% deviation.
An exact solution, taking into account the full hydrodynamics, is possible for an infinitely long flagellum in *two* spatial dimensions (where hydrodynamics is more long-ranged than in three dimensions) — corresponding to an infinite sheet with a propagating lateral wave with transverse oscillations in three dimensions. Here, the swimming velocity is obtained in pioneering work by G.I. Taylor to be [102]

$$v = \frac{1}{2}A^2\omega k\left(1 - \frac{19}{16}A^2k^2\right) \; . \tag{26}$$

This result confirms all qualitative features discussed above, but shows somewhat different numerical coefficients (which is in part due to the different dimensionality).
The sperm structure or beat pattern is typically not completely symmetric, but has some chirality. In this case, sperm swim on helical trajectories [103, 104]. For example, the helicity of the swimming trajectories is very pronounced for sea urchin sperm [105–107].
The same methods can be employed to calculate the swimming velocity of bacteria with rotating helical flagella [108].

4 Active Brownian Particles

The dynamics of microswimmers is governed by their self-propulsion, rotational and translational diffusion, and the interactions with other microswimmers and with obstacles and surfaces.

These interactions include both direct interactions (like volume exclusion or Coulomb repulsion) and hydrodynamic interactions. The importance of noise for the dynamics of microswimmers can be estimated by the Péclet number Pe, which compares advective and diffusive time scales. The self-advection time scale is L/v_0, where v_0 and L are the typical swimmer velocity and length, respectively, and the diffusive time scale $1/D_R$, with the rotational diffusion coefficient D_R. Hence, the (rotational) Péclet number is $Pe = v_0/(LD_R)$. For $Pe < 1$ the particle behaves like a passive Brownian particle, while for $Pe > 1$ the self-propulsion dominates. In cases where hydrodynamic interactions are not important, the system can be described by active Brownian particles.

4.1 Persistent Random Motion

Let us first consider the motion of a single active particle. Since the reorientation of the direction of motion is governed by the rotational diffusion coefficient D_R, the unit directional vector $\mathbf{n}$ behaves as a function of time t exactly as for a passive particle, i.e. $\langle \mathbf{n}(t)\cdot\mathbf{n}(0)\rangle = \exp(-t/\tau_R)$, where $\tau_R = 1/(2D_R)$ is the rotational diffusion time. In the limit that the particle moves essentially with constant (magnitude of the) velocity, i.e. $\mathbf{v}(t) = v_0\mathbf{n}(t)$, the particle position is $\mathbf{r}(t) = \mathbf{r}(0) + \int_0^t dt'\mathbf{v}(t')$. This implies

$$\Delta r^2 \equiv \langle(\mathbf{r}(t) - \mathbf{r}(0))^2\rangle = v_0^2 \int_0^t dt' \int_0^t dt'' \exp(-|t' - t''|/\tau_R) \tag{27}$$

so that

$$\Delta r^2 = 2v_0^2\tau_R^2\left[t/\tau_R + \exp(-t/\tau_R) - 1\right]. \tag{28}$$

Equation (28) shows a ballistic motion with $\Delta r^2 = v_0 2t^2$ for $t \ll \tau_R$, and a diffusive motion with $\Delta L^2 = (2v_0^2\tau_R)t$ for $t \gg \tau_R$. Thus, if we add (by hand) also the passive translational diffusion coefficient D_T, the effective translational diffusion coefficient becomes [7, 41]

$$D_{T,eff} = D_T + v_0^2/(6D_R), \tag{29}$$

which is much larger than D_T for $v_0 \gg \sqrt{D_T D_R}$.
The rotational and and translational diffusion coefficients for self-propelled Janus-colloids have been measured as a function of fuel concentration [41]. The translational diffusion is found to agree very well with $D_{T,eff}$ of Eq. (29).

4.2 Active Brownian Spheres near Surfaces

When active Brownian particles are confined between walls, self-propulsion leads to accumulation of particles near the wall, see Fig. 14. It is important to notice that spheres have no alignment interaction with the walls, so that the rotational diffusion is completely independent of the presence of the wall. Results of Brownian Dynamics simulations are presented in Fig. 15. They demonstrate that the (normalized) probability density $\rho(\Delta z)$ to find a particle at a distance Δz from the wall is strongly peaked close to the wall for $Pe \gtrsim 5$. Figure 15 shows the surface excess s as a function of the Péclet number $Pe = v_0/\sqrt{D_R D_T}$ for different channel widths. Note that s is a monotonically increasing function of Pe, and approaches unity for large Pe (complete adhesion).

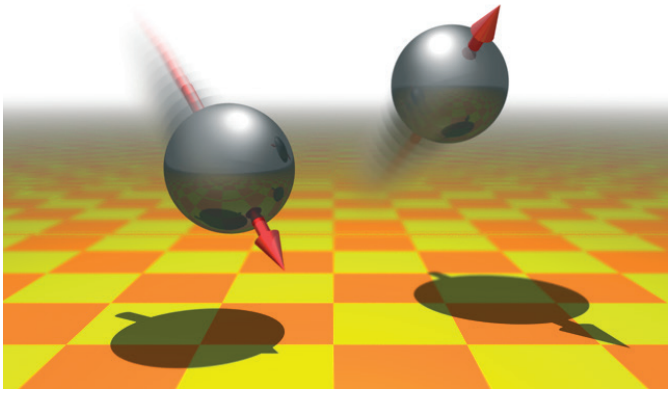


Fig. 14: *Schematic of active Brownian spheres moving near a surface. From Ref. [109].*

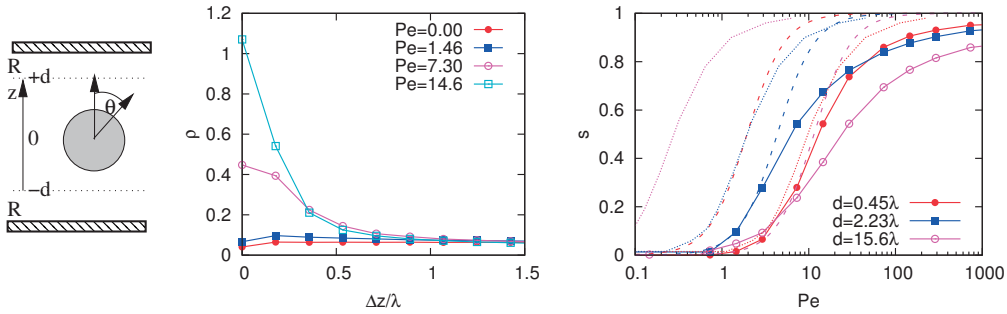


Fig. 15: *(***Left***) A self-propelled Brownian sphere of radius R is confined between two solid walls at $z = \pm(d + R)$. The orientation of the propulsion direction relative to the z-axis, is denoted by θ. (***Middle***) Probability density $\rho(\Delta z)$ to find a particle at a distance Δz from the wall. At zero Péclet number $Pe = v_0/\sqrt{D_R D_T}$, the probability density is uniform beyond the short range of the repulsive wall. With increasing Péclet number particles accumulate near the wall. Results are shown for a system with wall separation $d/\lambda = 15.6$, where $\lambda = \sqrt{D_T/D_R}$. (***Right***) Surface excess s as a function of Péclet number Pe for various wall separations d, as indicated. Results from analytic calculation for very narrow channels (dotted lines) and for small Péclet numbers (dashed lines) are also shown. All analytic expression have no adjustable parameters. From Ref. [110].*

The decoupling of the rotational degrees of freedom from translational motion allows for an analytic treatment via the Fokker-Planck equation [110]

$$\begin{aligned} \partial_t \rho(z,\theta,t) &= D_R \frac{1}{\sin(\theta)} \partial_\theta \left[\sin(\theta)\partial_\theta \rho(z,\theta,t)\right] \\ &- v_0 \cos(\theta)\partial_z \rho(z,\theta,t) + D_T \partial_z^2 \rho(z,\theta,t), \end{aligned} \tag{30}$$

where the angle $\theta = 0$ corresponds to particles oriented in the positive z-direction. This equation already demonstrates the main origin of surface accumulation. The rotational diffusion is independent of the spatial position, but particles are driven to one of the chamber walls depending on their orientation. Thus particles oriented toward the top, accumulate at the top wall, those pointing down, accumulate at the bottom wall. Less particles remain in the center. Solutions for small Péclet number and narrow channels are depicted in Fig. 15, and work well within their respective limits [110]. For example, the probability density $\rho(z,\theta)$ for narrow channels is found to be [110]

$$\rho(z,\theta) = \frac{Pe_d \rho_0 \cos(\theta)}{\pi \sinh(Pe_d \cos(\theta))} \exp[Pe_d \cos(\theta) z] \tag{31}$$

with $Pe_d = Pe\, d/\lambda$ and $\lambda = \sqrt{D_T/D_R}$.

4.3 Active Brownian Rods near Surfaces

In order to elucidate the effect of shape on the wall adhesion of a self-propelled particle due to steric interactions, it is interesting to consider the behavior of self-propelled Brownian rods near walls – again in the absence of any hydrodynamic interactions, but with Brownian noise [111, 112]. In this case, excluded-volume interactions favors parallel orientation along the wall, while the thermal noise leads to fluctuations of the rod orientation and thereby an effective repulsion from the wall. The competition of these two effects gives rise to an interesting adsorption behavior [111, 112].

Results of Brownian Dynamics simulations [111] are shown in Fig. 16. While passive rods are depleted from the surface (because their entropy is reduced near the surface due to restricted orientational fluctuations), active rods show an increased probability density near the surface, which grows with increasing propelling force f_t, see Fig. 16(left). In addition to the propelling force, the behavior of the rods strongly depends on the rod length L. The surface excess — the integrated probability density to find a rod near the surface compared to a uniform density distribution — is shown in Fig. 16(left). The results show that (i) short rods show no surface aggregation for any propelling force, and (ii) the surface excess initially increases with increasing f_t and L, but then saturates and becomes nearly independent of f_t and L for large propelling forces and rod lengths [111]. In comparison with active Brownian spheres, it is interesting to note that for rods, the surface excess never reaches unity. Thus, the elongated shape leads to a *reduction* of surface accumulation due to rapid alignment with the surface upon collision.

In order to understand the mechanism which is responsible for the effective surface adhesion of self-propelled rods, and to predict their behavior as a function of rod length, propulsive force, and wall separation, the analogy of the trajectories of self-propelled rods with the conformations of semi-flexible polymers can be exploited [111]. In the bulk, the rotational diffusion constant of a rod is $D_R \sim k_B T/(\eta L^3)$, compare Sec. 3.1, which implies a persistence length

$$\xi_p \sim v/D_R \sim \eta v l^3/k_B T \tag{32}$$

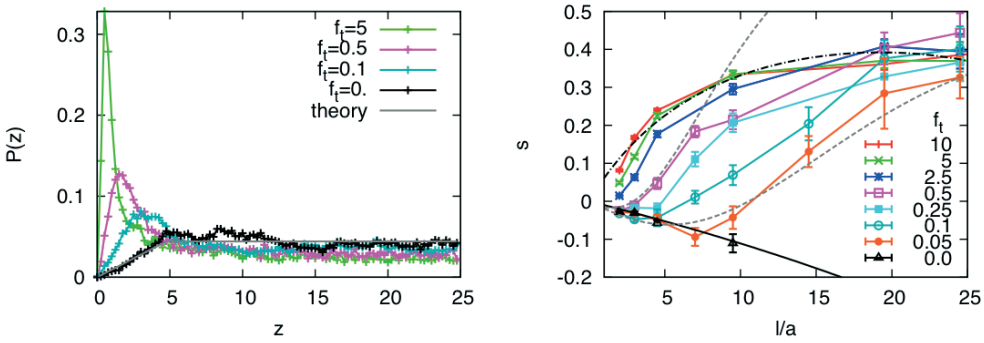


Fig. 16: *(Left) Probability density $P(z)$ as function of the distance z from the surface, for various propelling forces f_t. The rod length is $l \simeq 10a$ (where a is the rod diameter), the walls are located at $z = 0$ and $z = 5l$. A solid gray line shows the density profile of passive rods. (Right) Surface excess s as a function of scaled rod length l/a, for various propelling forces f_t, as indicated. The (black) dashed-dotted line is the scaling result in the ballistic regime (see Eq. (35)), the (gray) dashed lines are scaling results in the diffusive regime for $f_t = 0.5$ and $f_t = 0.05$. From Ref. [111].*

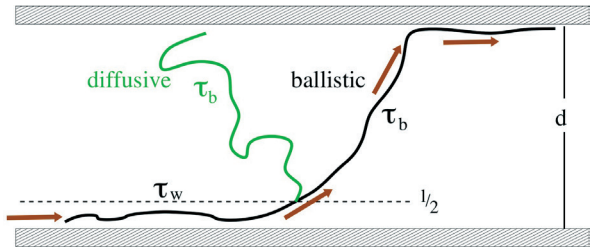


Fig. 17: *Schematic representation of the different regimes of rod motion, near a wall for time τ_w, and in the bulk for time τ_b, either in the ballistic or in the diffusive regime. From Ref. [111].*

of the trajectory. The probability p to find the self-propelled rod in a layer of thickness $L/2$ near the wall can be expressed as $p = \tau_w/(\tau_w + \tau_b)$, where τ_w is the time the rod remains within this layer and τ_b is the time it is located in the bulk (with $L/2 < z < d - L/2$).

To estimate τ_w, let us consider a rod, which at time $t = 0$ is oriented parallel to the wall, and located very close to the wall with $0 < z \ll L/2$. As the rod moves forward, it is reflected when is hits the wall, and is thereby constrained to the positive half-space $z > 0$, see Fig. 17. This situation is very similar to a semi-flexible polymer, which is fixed at one end near the wall with tangent vector parallel to the wall; its bending rigidity κ is determined by the persistence length, $\xi_p = \kappa/k_BT$. In this case, the distance of the polymers from the wall increases as $\langle z\rangle \sim (k_BT/\kappa)^{1/2}x^{3/2}$ and the orientation angle as $\langle\theta\rangle \sim (k_BT/\kappa)^{1/2}x^{1/2}$, where $x = vt$ is the distance traveled parallel to the wall [113, 114]. The condition $\langle z\rangle = l/2$ at $t = \tau_w$ then implies [111]

$$\tau_w \sim \frac{1}{v}\left(l^2\xi_p\right)^{1/3} \sim \left(\frac{\eta}{k_BT}\right)^{1/3} l^{5/3}\, v^{-2/3}. \tag{33}$$

For the time τ_b for the rod to stay in the bulk fluid, we have to distinguish two regimes. In the

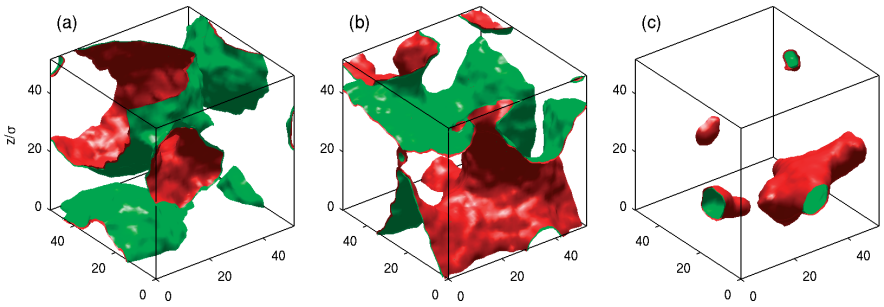


Fig. 18: *Snapshots of the gas-liquid interfaces of a three-dimensional system of active Brownian spheres (a,b) just above the clustering transition (volume fractions $\phi = 0.3375$ and $\phi = 0.4$) and (c) at a high density ($\phi = 0.6$). The red and green iso-density surfaces correspond to $\phi_l = 0.54$ and $\phi_l = 0.5$, respectively. From Ref. [116].*

ballistic regime, with $\xi_p \gg d$, the rod travels essentially on a straight line between the walls, see Fig. 17. In this case, the bulk time is given by $\tau_b \sim v^{-1} d/\sin(\theta)$, where θ is the angle of the rod with the surface when it leaves the wall layer of thickness $l/2$. The polymer analogy explained above implies $\langle\theta\rangle \sim (l/\xi_p)^{1/3}$ for $\theta \ll 1$, so that [111]

$$\tau_b \sim \frac{d}{v}\left(\frac{\xi_p}{l}\right)^{1/3} \sim \left(\frac{\eta}{k_B T}\right)^{1/3} d\, l^{2/3}\, v^{-2/3}. \tag{34}$$

Thus, the scaling arguments predict in the ballistic regime the probability

$$p = l/(l + a_B d) \tag{35}$$

to find the rod in the wall layers, with a constant a_B which has to be determined numerically. Note that this expression is independent of the velocity v, because both time scales τ_w and τ_b depend on v in the same way. Thus, scaling theory explains the saturation of the excess surface density with increasing propelling force, as shown in Fig. 16(right).
For a system of many self-propelled Brownian rods between two walls in two dimensions, the rods moving along the walls in opposite directions block each other and lead to the formation of "hedgehog-like" clusters [115].

4.4 Collective Aggregation of Active Brownian Spheres

Active Brownian spheres exhibit a fascinatingly rich collective behavior [116, 117]. At low densities and swimming velocities, the particles in three spatial dimensions show a gas-like behavior comparable to passive particles. However, above a certain Péclet number and density, they phase separate into a dense fluid and a dilute gas phase, see Fig. 18. Here, the local density of the fluid phase $\phi_l = 0.62$ is considerably higher than the critical density $\phi_l = 0.56$ for glass formation of passive particles in three-dimensional space. At a given Péclet number, the morphological character of the two-phase region changes with increasing density from a fluid droplet over a bicontinuous structure to a gas-phase droplet embedded in a fluid phase [116].

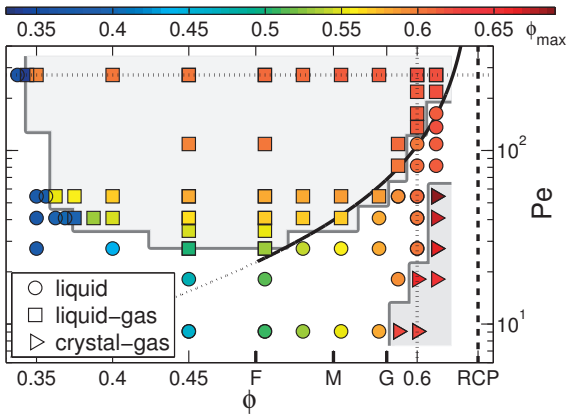


Fig. 19: *Phase diagram of an suspension of active Brownian spheres. Symbols denote the homogeneous liquid phase (○), the gas-liquid coexistence (□) and the crystal-gas coexistence (▷). The equilibrium transition points of hard-spheres for freezing ($\phi_F = 0.494$), melting ($\phi_M = 0.545$), glass-transition point ($\phi_G \approx 0.58$), and random close packing ($\phi_{RCP} \approx 0.64$), are indicated by F, M, G, RCP, respectively. From Ref. [116].*

In contrast to two-dimensional systems of active Brownian disks [54, 91, 92], where crystalline clusters are formed, the dense phase is fluid-like up to rather high densities close to random-closed packing, see Fig. 19. Here, the glass-transition density shifts to higher average packing fractions [116, 118].

Particularly remarkable is the dynamical behavior of the active Brownian spheres in the dense phase, where they exhibit jets, swirls, and turbulence, see Fig. 20, similar to experimental observations in bacterial colonies. This is surprising, because there is no alignment mechanism between the moving directions of individual particles. Activity and excluded volume interactions suffice to generate highly complex dynamical patterns. A possible explanation for these patterns is a sorting of particles at interfaces [116]. At an interface, particles accumulate with their force direction preferentially toward the dense phase. It then takes a while for their force direction to reorient due to rotational diffusion such that they can move away from the interface. This waiting time depends also on the interface curvature. Therefore, particles remain for a longer time at concave than in convex regions, which leads to an accumulation of particles with similar orientation. Such oriented clusters could then provide the driving force for larger jets and swirls.

The collective motion of rod-like microswimmers has also been studied intensively [7, 119–124]. In this case, star-like immobile clusters of motile clusters with nematic order can form. A detailed discussion goes beyond the scope of this Chapter.

5 Hydrodynamic Synchronization

In nature, the density of flagella, cilia, and various kinds of microswimmers can sometimes be very high. For example, in mammalian reproduction, the average number of *sperm* per ejaculate is tens to hundreds of millions, so that the average distance between sperm is on the scale of ten micrometers — comparable to the length of their flagellum — so that interactions

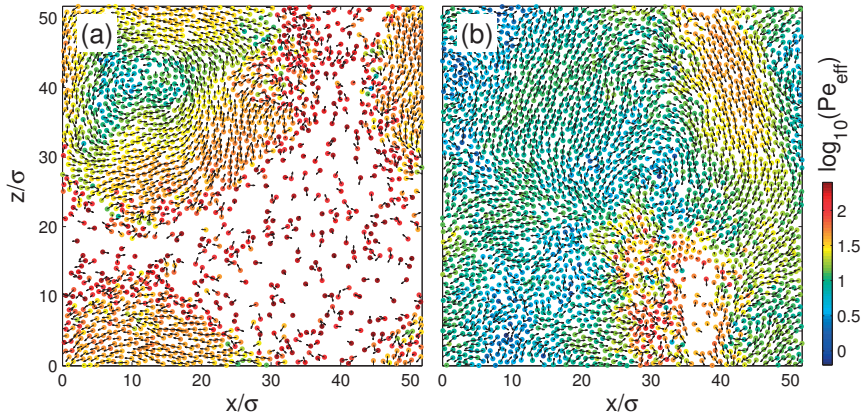


Fig. 20: *Collective motion of active Brownian spheres, in the steady state at Péclet number* $Pe = 272$*. Snapshots (slices of thickness* σ*) of a system (a) just above the clustering transition (*$\phi = 0.3375$*) and (b) at a high concentration (*$\phi = 0.6$*). Arrows indicate the direction of the displacements over a short lag time. The magnitude is color-coded and is expressed by the effective Péclet number* Pe_{eff}*. From Ref. [116].*

between them are not negligible. In recent years, experiments [125–130] have revealed an interesting cooperative behavior of sperm at high concentration, e.g. the distinctive aggregations or 'trains' of hundreds of wood-mouse sperm [128, 129], or the vortex arrays of swimming sea urchin sperm on a substrate [130]. Vortex arrays of curved flagella have also been obtained in simulations [131]. *Cilia* densely cover the surfaces of the protozoa Paramecium [132] and Opalina [133, 134] and of the green algae *Volvox* [40, 135], and line the airways in the human body. Lophotrichous and peritrichous *bacteria* have multiple flagella located at the same spot and located randomly, respectively, on the bacteria surface. Their run-and-tumble motion requires the synchronization, bundling and unbundling of these flagella [136–141]. In all these cases, hydrodynamic interactions play an important role.

5.1 Synchronization of Cilia in Metachronal Waves

Motile cilia on the surface of a cell or microorganism perform an active whip-like motion, which propels the fluid along the surface of cells and tissues. As explained in Sec. 1, this beat consists of a fast power stroke in which the cilium has an elongated shape, and a slower recovery stroke in which the cilium is curved and closer to the cell surface. The beat of different cilia is usually not random, but strongly synchronized in a wave-like pattern which is called a *metachronal wave* (MCW) [133].

Theoretical approaches to investigate hydrodynamic interactions between cilia and the formation of metachronal waves fall in three categories: (i) highly simplified model systems, designed to elucidate the mechanism of hydrodynamic synchronization of many active agents [12, 135, 142–147], (ii) models of an actively driven semi-flexible filament, which mimic the beat of a real cilium [148–152], and (iii) models of a filament, with a beat shape obtained from maximizing the pumping efficiency [153, 154].

The second class of models consists of semi-flexible filaments, which are deformed actively by

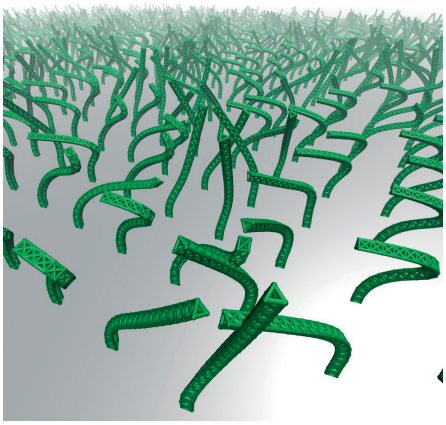


Fig. 21: *Simulation snapshot of cilia conformations during the formation of metachronal waves. From Ref. [152].*

internal forces to reproduce the power and recovery strokes of real cilia, and can react to the flow field generated by their neighbors [148–151, 155]. The studies of such models have been restricted so far mainly to effectively one-dimensional chains of cilia [148–151], or to two-dimensional arrays of a small number of cilia [155]. They provide an indication of metachronal coordination, but the considered systems were too small or the time evolution too short to allow any prediction of MCW properties. This limitation has been overcome very recently by a large-scale, mesoscopic hydrodynamics simulation of arrays of up to 60×60 cilia, which are modeled as active filaments with geometric triggers for the switching between power and recovery strokes [152], see Fig. 21. This model provides a very clear evidence for the emergence of metachronal waves, and allows a detailed study of wave and transport properties [152].
Simulation results for the beat period, the fluid velocity above the cilia array, and the transport efficiency are compared in Fig. 22 for cilia arrays with either metachronal coordination or a completely synchronous beat [152]. The first surprising result is that the time required for a single beat actually increases in the presence of metachronal waves (MCWs), see Fig. 22(a). This can be understood be the larger hydrodynamic resistance when cilia are beating against each other in a MCW. However, the faster beating in synchronous beating is to little avail, since the fluid above is just pushed back and forth, which results in a much smaller net fluid transport velocity compared to MCWs, see Fig. 22(b). This is not only inefficient in the sense of generating only slow fluid transport, but also from an energetic point of view, see Fig. 22(c), because also the energy efficiency

$$\epsilon = \frac{16}{3}\frac{\eta v^2 d_c^2}{L_c P_c} \tag{36}$$

(where η is the fluid viscosity, v the average fluid velocity, d_c and L_c the cilium separation and length, respectively, and P_c the energy consumption per unit time) is much lower for synchronous beating (because also oscillating a viscous fluid generates dissipation) than for MCWs. This effect can be quite large, reaching almost an order of magnitude for the optimal cilia spacing.
From the experimental side, a detailed quantitative investigation of metachronal waves and their properties has been performed recently in Ref. [135] for *Volvox carteri*, whose large size and

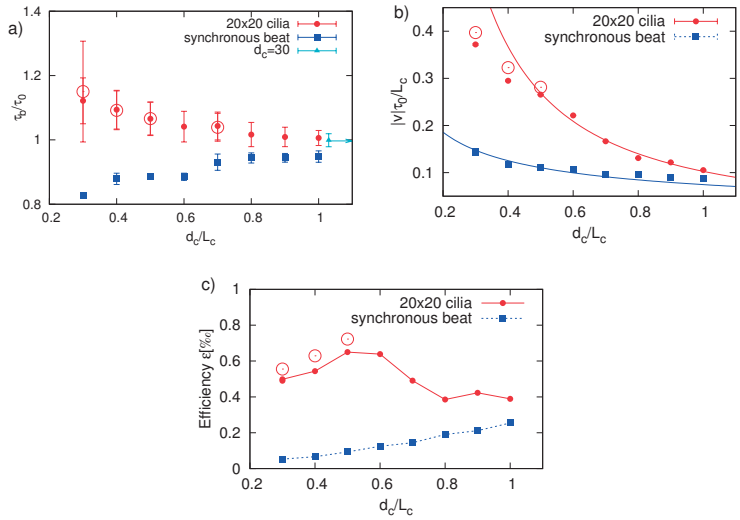


Fig. 22: *Mesoscale hydrodynamics simulations show that metachronal coordination greatly increases swimming velocity. (a) Beat period τ_b versus cilia spacing d_c. (b) Average fluid velocity v. Solid lines are fits to power laws ${d_c}^{-\nu}$ (using data for $d_c \geq 5$ with 20×20 cilia). (c) Dimensionless efficiency ϵ. In all figures, results are shown for MCWs in arrays of 20×20 cilia (red bullets) and in arrays of 60×60 cilia (red circles), as well as for synchronously beating cilia (blue squares). Error bars denote standard deviation. From Ref. [152].*

ease of visualization make it an ideal model organism for such studies. In particular, beat frequencies, decay times, and wave numbers have been determined. Interestingly, the decay time is found to be just a few times the beat period.

6 Summary and Conclusions

The swimming behavior of biological and synthetic microswimmers is an exciting field of current research. Interesting aspects of swimming are (i) the derivation of the velocity of a swimmer from the underlying propulsion mechanism, (ii) the prediction of the dynamics of two swimmers interacting with each other via static (conservative) forces and hydrodynamic interactions, (iii) the determination of the swimming behavior near walls and in confined geometry, and (iv) the prediction of the collective behavior of many interacting swimmers.

The understanding of the behavior of the "classical" biological microswimmers (such as sperm, bacteria and algae) has increased a lot, and many novel microswimmers have been designed in recent years and decades. This opens the way for a more detailed understanding and many new applications of microswimmers. What is universal about the swimming behavior (and can thus be described by the hydrodynamic far-field approximation), and what is specific for a certain class of swimmers (like the synchronization of different wave forms of flagella)? How can synthetic microswimmers be designed to react to external stimuli and find their targets? How can the "biological complications" be incorporated into theoretical models? The investigations of these and other questions make microswimmers also a exciting research topic in the future.

Appendices

A Oseen Tensor of Low-Reynolds-Number Hydrodynamics

The Stokes equation

$$-\nabla p(\mathbf{r}) + \eta \nabla^2 \mathbf{u}(\mathbf{r}) + \mathbf{f}(\mathbf{r}) = 0 \tag{37}$$

for the hydrodynamic velocity field $\mathbf{u}(\mathbf{r})$ and the pressure field $p(\mathbf{r})$ of an incompressible fluid [with $\nabla \cdot \mathbf{u}(\mathbf{r}) = 0$] can be solved explicitly for a body-force field $\mathbf{f}(\mathbf{r})$ by Fourier transformation, as explained in detail in Chapter B.3 (Winkler). With

$$p(\mathbf{k}) = \int d^3 r e^{i\mathbf{k}\cdot\mathbf{r}} p(\mathbf{r}) \tag{38}$$

and similarly $\mathbf{u}(\mathbf{k})$, etc., the Stokes equation and the incompressibility condition become

$$-i\mathbf{k}p(\mathbf{k}) - \eta k^2 \mathbf{u}(\mathbf{k}) + \mathbf{f}(\mathbf{k}) = 0 \tag{39}$$

$$\mathbf{k} \cdot \mathbf{u}(\mathbf{k}) = 0 \, . \tag{40}$$

The multiplication of Eq. (39) with $\mathbf{k}$ then implies

$$p(\mathbf{k}) = -i\mathbf{k} \cdot \mathbf{f}(\mathbf{k})/k^2 \, . \tag{41}$$

Insertion of this result in Eq. (39) gives

$$-\mathbf{k}(\mathbf{k} \cdot \mathbf{f}(\mathbf{k}))/k^2 - \eta k^2 \mathbf{u}(\mathbf{k}) + \mathbf{f}(\mathbf{k}) = 0 \tag{42}$$

so that

$$u_\alpha(\mathbf{k}) = \frac{1}{\eta k^2} \sum_\beta \left[\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2} \right] f_\beta(\mathbf{k}) \, . \tag{43}$$

Fourier transformation back to real space then yields (convolution theorem)

$$\mathbf{u}(\mathbf{r}) = \int d^3 r' \mathbf{H}(\mathbf{r} - \mathbf{r}') \cdot \mathbf{f}(\mathbf{r}') \tag{44}$$

with the Oseen tensor

$$H_{\alpha\beta}(\mathbf{r}) = \int \frac{d^3 k}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{r}} \frac{1}{\eta k^2} \left[\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2} \right] \, . \tag{45}$$

Because the tensor $\mathbf{H}(\mathbf{r})$ only depends on the vector $\mathbf{r}$, it can be written in the form

$$H_{\alpha\beta}(\mathbf{r}) = A(r)\delta_{\alpha\beta} + B(r) r_\alpha r_\beta / r^2 \, ; \tag{46}$$

Thus,

$$\sum_\alpha H_{\alpha\alpha}(\mathbf{r}) = 3A(r) + B(r) \tag{47}$$

$$\sum_{\alpha\beta} H_{\alpha\beta}(\mathbf{r}) r_\alpha r_\beta / r^2 = A(r) + B(r) \, . \tag{48}$$

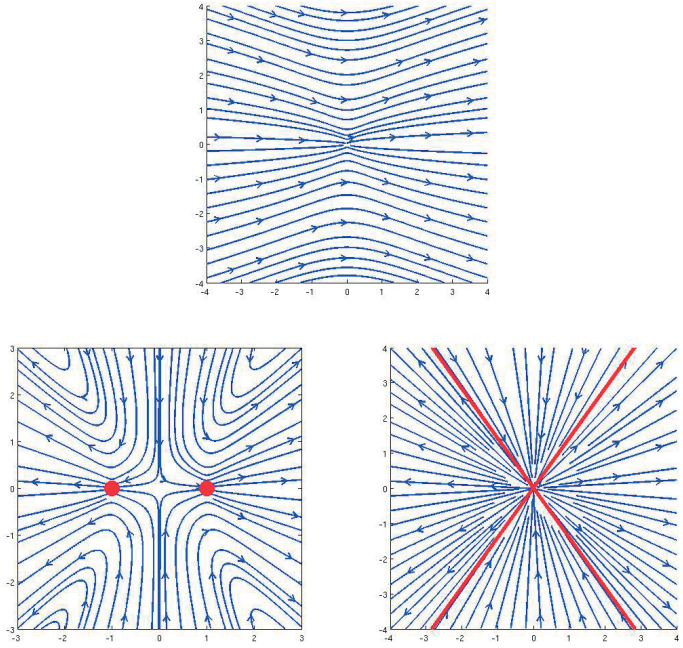


Fig. 23: *Flow lines of hydrodynamic monopole and dipole, oriented in the horizontal direction. (a) Far-field of the monopole, as given by Eq. (51). (b) Near-field of the dipole, as given by Eq. (53). The two force centers are marked by (red) bullets. (c) Far-field of the dipole, as given by Eq. (54). The separatrix between the inflow and outflow regions is shown by thick red lines.*

An explicit calculation using Eq. (45) provides

$$\sum_{\alpha} H_{\alpha\alpha}(\mathbf{r}) = \frac{1}{2\pi\eta r} \quad ; \quad \sum_{\alpha\beta} H_{\alpha\beta}(\mathbf{r}) r_\alpha r_\beta / r^2 = \frac{1}{4\pi\eta r} \; . \tag{49}$$

Thus, the Oseen tensor finally reads

$$H_{\alpha\beta}(\mathbf{r}) = \frac{1}{8\pi\eta r}\left[\delta_{\alpha\beta} + \frac{r_\alpha r_\beta}{r^2}\right] \; . \tag{50}$$

The Oseen tensor shows that the hydrodynamic interaction is very long ranged, with a $1/r$ decay like the Coulomb potential. Equations (44) and (50) imply that the fluid velocity field for a point force at the origin in direction $\hat{\mathbf{e}}$, $\mathbf{f}(\mathbf{r}) = \hat{\mathbf{e}}\delta(\mathbf{r})$, is given by

$$\mathbf{u}(\mathbf{r}) = \frac{1}{8\pi\eta r}\left[\hat{\mathbf{e}} + \frac{(\mathbf{r}\cdot\hat{\mathbf{e}})\mathbf{r}}{r^2}\right] \; . \tag{51}$$

The streamlines of a force monopole is shown in Fig. 23(a)

B Hydrodynamic Force Dipoles

The flow field of a hydrodynamic force dipole, with opposite forces of equal magnitude at $\mathbf{r} = \pm\mathbf{r}_0$ and direction $\hat{\mathbf{e}} = \mathbf{r}_0/r_0$,

$$\mathbf{f}_1(\mathbf{r}) = f_0\hat{\mathbf{e}}\delta(\mathbf{r}-\mathbf{r}_0) \quad ; \quad \mathbf{f}_2(\mathbf{r}) = -f_0\hat{\mathbf{e}}\delta(\mathbf{r}+\mathbf{r}_0) \tag{52}$$

is obtained from Eq. (51) to be

$$\begin{aligned}\mathbf{u}(\mathbf{r}) &= \frac{1}{8\pi\eta\sqrt{(\mathbf{r}-\mathbf{r}_0)^2}}\left[\hat{\mathbf{e}} + \frac{((\mathbf{r}-\mathbf{r}_0)\cdot\hat{\mathbf{e}})(\mathbf{r}-\mathbf{r}_0)}{(\mathbf{r}-\mathbf{r}_0)^2}\right] \\ &- \frac{1}{8\pi\eta\sqrt{(\mathbf{r}+\mathbf{r}_0)^2}}\left[\hat{\mathbf{e}} + \frac{((\mathbf{r}+\mathbf{r}_0)\cdot\hat{\mathbf{e}})(\mathbf{r}+\mathbf{r}_0)}{(\mathbf{r}+\mathbf{r}_0)^2}\right]\end{aligned} \tag{53}$$

The flow lines in the near-field of this force dipole are shown in Fig. 23(b).
An expansion to leading order in $\mathbf{r}_0/|\mathbf{r}|$ yields

$$\mathbf{u}(\mathbf{r}) = \frac{P}{8\pi\eta r^3}\left[-1 + 3\frac{(\mathbf{r}\cdot\hat{\mathbf{e}})^2}{r^2}\right]\mathbf{r} \tag{54}$$

where $P = 2f_0r_0$ is the dipole strength. Note that the flow field of a force dipole decays as $1/r^2$ from the center of the dipole, faster than the force monopole of Eq. (51). The flow lines of a hydrodynamic dipole are shown in Fig. 23(c). There are two inflow and two outflow regions in the xy-projection, which are determined by the separatrix $y = \pm\sqrt{2}x$. In three dimensions, the outflow region is of course a cone.

C Hydrodynamics near Planar Surfaces

An exact solution of the Stokes equation near a planar surface with no-slip boundary conditions (i.e. $\mathbf{u}(\mathbf{r}) = 0$ at the surface) is possible. The solution is called the Blake tensor [156]. Instead of the exact solution, we consider here a simple approximation for a dipole near a planar wall, oriented parallel to the wall, by employing the method of image charges known from electrostatics, i.e. we write for a wall at $z = 0$,

$$\mathbf{u}_{wall}(\mathbf{r}-\mathbf{r}_0) = \mathbf{u}_{dipole}(\mathbf{r}-\mathbf{r}_0;\hat{\mathbf{e}}) + \mathbf{u}_{dipole}(\mathbf{r}-\mathbf{r}_1;\hat{\mathbf{e}}') \tag{55}$$

where $\mathbf{r}_0 = (x_0, y_0, z_0)$ and $\mathbf{r}_1 = (x_0, y_0, -z_0)$, with $z_0 > 0$, and $\hat{\mathbf{e}}'$ is the mirror image of $\hat{\mathbf{e}}$ with respect to the $z = 0$ plane . This implies that at $z = 0$ the velocity field u_z perpendicular to the surface vanishes identically, $u_z \equiv 0$, but the no-slip boundary condition is *not* satisfied. The dipole experiences a force near the surface, which is determined by the hydrodynamic interaction between the dipole and the image charge. It is given by the z-component of the flow field of the image charge at the location of the dipole, so that

$$v_z(z_0) = -\frac{P}{32\pi\eta z_0^2}\left[1 - 3(\hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_z)^2\right] \ . \tag{56}$$

because $(\hat{\mathbf{e}}'\cdot\hat{\mathbf{e}}_z)^2 = (\hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_z)^2$. This result shows that the hydrodynamic force is attractive to the wall, and that it decays as the dipole flow field with the squared distance from the wall. Eq. (56) is in agreement with the exact result (9) for no-slip boundary conditions, except that the numerical prefactor in Eq. (56) is smaller by a factor $2/3$.

References

[1] E. M. Purcell, Am. J. Phys. **45**, 3 (1977).

[2] J. Toner, Y. Tu, and S. Ramaswamy, Ann. Phys. (N.Y.) **318**, 170 (2005).

[3] S. Ramaswamy, Ann. Rev. Condens. Matter Phys. **1**, 323 (2010).

[4] T. Vicsek and A. Zafeiris, Phys. Rep. **517**, 71 (2012).

[5] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Rev. Mod. Phys. **85**, 1143 (2013).

[6] D. Saintillan and M. J. Shelley, C. R. Physique **14**, 497 (2013).

[7] J. Elgeti, R. G. Winkler, and G. Gompper, Rep. Prog. Phys., to appear (2015); arXiv:1412.2692.

[8] E. Lauga and T. R. Powers, Rep. Prog. Phys. **72**, 096601 (2009).

[9] T. Ishikawa, J. R. Soc. Interface **6**, 815 (2009).

[10] G. Gompper, in *Lecture Notes of 42nd IFF Springschool: Macromolecular Systems in Soft and Living Matter*, edited by J. Dhont, G. Gompper, P. Lang, D. Richter, M. Ripoll, D. Willbold, and R. Zorn (Forschungszentrum Jülich, Jülich, 2011), vol. 20 of *Key Technologies*.

[11] D. L. Koch and G. Subramanian, Ann. Rev. Fluid Mech. **43**, 637 (2011).

[12] R. Golestanian, J. M. Yeomans, and N. Uchida, Soft Matter **7**, 3074 (2011).

[13] R. M. Harshey, Annu. Rev. Microbiol. **57**, 249 (2003).

[14] M. E. Cates, Rep. Prog. Phys. **75**, 042601 (2012).

[15] L. Alvarez, B. M. Friedrich, G. Gompper, and U. B. Kaupp, Trends Cell Biol. **24**, 198 (2014).

[16] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, Eur. Phys. J. Special Topics **1**, 202 (2012).

[17] G. A. Ozin, I. Manners, S. Fournier-Bidoz, and A. Arsenault, Adv. Mater. **17**, 3011 (2005).

[18] S. Sengupta, M. E. Ibele, and A. Sen, Angew. Chem. Int. Ed. **51**, 8434 (2012).

[19] S. J. Ebbens and J. R. Howse, Soft Matter **6**, 726 (2010).

[20] H. C. Berg, *E. coli in Motion* (Springer, New York, 2004).

[21] C. Brennen and H. Winet, Ann. Rev. Fluid Mech. **9**, 339 (1977).

[22] H. C. Berg, Annu. Rev. Biochem. **72**, 19 (2003).

[23] D. Jones, New Scientist **197**, 40 (2008).

[24] N. C. Darnton, L. Turner, S. Rojevsky, and H. C. Berg, J. Bacteriol. **189**, 1756 (2007).

[25] Http://en.wikipedia.org/wiki/Spermatozoon.

[26] Http://www.tutorvista.com/topic/cilia-and-flagella.

[27] B. M. Friedrich, I. H. Riedel-Kruse, J. Howard, and F. Jülicher, J. Exp. Biol. **213**, 1226 (2010).

[28] S. L. Tamm and G. A. Horridge, Proc. R. Soc. Lond. B **175**, 219 (1970).

[29] D. R. Brumley, K. Y. Wan, M. Polin, and R. E. Goldstein, eLife **3**, e02750 (2014).

[30] C. B. Lindemann, J. Theor. Biol. **168**, 175 (1994).

[31] C. B. Lindemann and K. A. Lesich, J. Cell Sci. **123**, 519 (2010).

[32] C. B. Lindemann, Biophys. J. **107**, 1487 (2014).

[33] F. Jülicher and J. Prost, Phys. Rev. Lett. **78**, 4510 (1997).

[34] S. Camalet, F. Jülicher, and J. Prost, Phys. Rev. Lett. **82**, 1590 (1999).

[35] Http://en.wikipedia.org/wiki/Chlamydomonas_reinhardtii.

[36] K. Drescher, R. E. Goldstein, N. Michel, M. Polin, and I. Tuval, Phys. Rev. Lett. **105**, 168101 (2010).

[37] J. S. Guasto, K. A. Johnson, and J. P. Gollub, Phys. Rev. Lett. **105**, 168102 (2010).

[38] Http://remf.dartmouth.edu/images/algaeSEM.

[39] Http://en.wikipedia.org/wiki/Volvox.

[40] K. Drescher, K. C. Leptos, I. Tuval, T. Ishikawa, T. J. Pedley, and R. E. Goldstein, Phys. Rev. Lett. **102**, 168101 (2009).

[41] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian, Phys. Rev. Lett. **99**, 048102 (2007).

[42] A. Erbe, M. Zientara, L. Baraban, C. Kreidler, and P. Leiderer, J. Phys. Condens. Matter **20**, 404215 (2008).

[43] R. Dreyfus, J. Baudry, M. L. Roper, M. Fermigier, H. A. Stone, and J. Bibette, Nature **437**, 862 (2005).

[44] B. J. Williams, S. V. Anand, J. Rajagopalan, and M. T. A. Saif, Nat. Commun. **5**, 3081 (2014).

[45] W. F. Paxton, K. C. Kistler, C. C. Olmeda, A. Sen, S. K. St. Angelo, Y. Cao, T. E. Mallouk, P. E. Lammert, and V. H. Crespi, J. Am. Chem. Soc. **126**, 13424 (2004).

[46] A. Fournier-Bidoz, A. C. Arsenault, I. Manners, and G. A. Ozin, Chem. Commun., pp. 441–443 (2005).

[47] R. Golestanian, T. B. Liverpool, and A. Ajdari, Phys. Rev. Lett. **94**, 220801 (2005).

[48] G. Rückner and R. Kapral, Phys. Rev. Lett. **98**, 150603 (2007).

[49] R. Golestanian, Phys. Rev. Lett. **102**, 188305 (2009).

[50] M. N. Popescu, S. Dietrich, and G. Oshanin, J. Chem. Phys. **130**, 194702 (2009).

[51] M. N. Popescu, S. Dietrich, M. Tasinkevych, and J. Ralston, Eur. Phys. J. E **31**, 351 (2010).

[52] S. Thakur and R. Kapral, J. Chem. Phys. **135**, 024509 (2011).

[53] B. Sabass and U. Seifert, J. Chem. Phys. **136**, 064508 (2012).

[54] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin, Science **339**, 936 (2013).

[55] H.-R. Jiang, N. Yoshinaga, and M. Sano, Phys. Rev. Lett. **105**, 268302 (2010).

[56] M. Yang and M. Ripoll, Phys. Rev. E **84**, 061401 (2011).

[57] T. Bickel, A. Majee, and A. Würger, Phys. Rev. E **88**, 012301 (2013).

[58] P. de Buyl and R. Kapral, Nanoscale **5**, 1337 (2013).

[59] G. Volpe, I. Buttinoni, D. Vogt, H. J. Kümmerer, and C. Bechinger, Soft Matter **7**, 8810 (2011).

[60] I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, and C. Bechinger, J. Phys.: Condens. Matter **24**, 284129 (2012).

[61] Y. W. Kim and R. R. Netz, Phys. Rev. Lett. **96**, 158101 (2006).

[62] E. M. Gauger, M. T. Downton, and H. Stark, Eur. Phys. J. E **28**, 231 (2009).

[63] T. Sanchez, D. Welch, D. Nicastro, and Z. Dogic, Science **333**, 456 (2011).

[64] S. Sareh, J. Rossiter, A. Conn, K. Drescher, and R. E. Goldstein, J. R. Soc. Interface **10**, 20120666 (2012).

[65] A. Keißner and C. Brücker, Soft Matter **8**, 5342 (2012).

[66] J. Branscomb and A. Alexeev, Soft Matter **6**, 4066 (2010).

[67] H. Masoud and A. Alexeev, Soft Matter **7**, 8702 (2011).

[68] M. Vilfan, A. Potočnik, B. Kavčič, N. Osterman, I. Poberaj, A. Vilfan, and D. Babič, Proc. Natl. Acad. Sci. USA **107**, 1844 (2010).

[69] A. Najafi and R. Golestanian, Phys. Rev. E **69**, 062901 (2004).

[70] R. Golestanian, Eur. Phys. J. E **25**, 1 (2008).

[71] G. P. Alexander, C. M. Pooley, and J. M. Yeomans, Phys. Rev. E **78**, 045302(R) (2008).

[72] M. J. Lighthill, Commun. Pure Appl. Math. **5**, 109 (1952).

[73] J. R. Blake, J. Fluid Mech. **46**, 199 (1971).

[74] T. Ishikawa, M. P. Simmonds, and T. J. Pedley, J. Fluid Mech. **568**, 119 (2006).

[75] T. Ishikawa, G. Sekiya, Y. Imai, and T. Yamaguchi, Biophys. J. **93**, 2217 (2007).

[76] I. O. Götze and G. Gompper, Phys. Rev. E **82**, 041921 (2010).

[77] T. Ishikawa and T. J. Pedley, Phys. Rev. Lett. **100**, 088103 (2008).

[78] A. Zöttl and H. Stark, Phys. Rev. Lett. **112**, 118101 (2014).

[79] B. A. Grzybowsk, H. A. Stone, and G. M. Whitesides, Nature **405**, 1033 (2000).

[80] I. Llopis and I. Pagonabarraga, Eur. Phys. J. E **26**, 103 (2008).

[81] P. Tierno, R. Golestanian, I. Pagonabarraga, and F. Sagues, Phys. Rev. Lett. **101**, 218304 (2008).

[82] S. Bleil, D. W. M. Marr, and C. Bechinger, Appl. Phys. Lett. **88**, 263515 (2006).

[83] I. O. Götze and G. Gompper, EPL **92**, 64003 (2010).

[84] I. O. Götze and G. Gompper, Phys. Rev. E **84**, 041921 (2011).

[85] E. Ben-Jacob, I. Cohen, and H. Levine, Adv. Phys. **49**(4), 395 (2000).

[86] A. Sokolov, I. S. Aranson, J. O. Kessler, and R. E. Goldstein, Phys. Rev. Lett. **98**, 158102 (2007).

[87] F. Peruani, J. Starruss, V. Jakovljevic, L. Søgaard-Andersen, A. Deutsch, and M. Bär, Phys. Rev. Lett. **108**, 098102 (2012).

[88] X. Chen, X. Dong, A. Be'er, H. L. Swinney, and H. P. Zhang, Phys. Rev. Lett. **108**, 148101 (2012).

[89] J. Gachelin, G. Miño, H. Berthet, A. Lindner, A. Rousselet, and E. Clément, Phys. Rev. Lett. **110**, 268103 (2013).

[90] M. Ibele, T. E. Mallouk, and A. Sen, Angew. Chem. Int. Ed. Engl. **48**, 3308 (2009).

[91] I. Theurkauff, C. Cottin-Bizonne, J. Palacci, C. Ybert, and L. Bocquet, Phys. Rev. Lett. **108**, 268303 (2012).

[92] I. Buttinoni, J. Bialké, F. Kümmel, H. Löwen, C. Bechinger, and T. Speck, Phys. Rev. Lett. **110**, 238301 (2013).

[93] W. Wang, W. Duan, A. Sen, and T. E. Mallouk, Proc. Natl. Acad. Sci. USA **110**, 17744 (2013).

[94] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Proc. Natl. Acad. Sci. USA **107**, 11865 (2010).

[95] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Phys. Rev. Lett. **75**, 1226 (1995).

[96] K. Drescher, J. Dunkel, L. H. Cisneros, S. Ganguly, and R. E. Goldstein, Proc. Natl. Acad. Sci. USA **108**, 10940 (2011).

[97] A. P. Berke, L. Turner, H. C. Berg, and E. Lauga, Phys. Rev. Lett. **101**, 038102 (2008).

[98] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).

[99] J. K. G. Dhont, *An Introduction to Dynamics of Colloids* (Elsevier, Amsterdam, 1996).

[100] J. Gray and G. J. Hancock, J. Exp. Biol. **32**, 802 (1955).

[101] J. J. L. Higdon, J. Fluid Mech. **90**, 685 (1979).

[102] G. I. Taylor, Proc. Roy. Soc. Lond. A **209**, 447 (1951).

[103] H. C. Crenshaw, Biophys. J. **56**, 1029 (1989).

[104] J. Elgeti, U. B. Kaupp, and G. Gompper, Biophys. J. **99**, 1018 (2010).

[105] H. C. Crenshaw, Am. Zool. **36**, 608 (1996).

[106] U. B. Kaupp, J. Solzin, E. Hildebrand, J. E. Brown, A. Helbig, V. Hagen, M. Beyermann, F. Pampaloni, and I. Weyand, Nat. Cell Biol. **5**, 109 (2003).

[107] M. Böhmer, Q. Van, I. Weyand, V. Hagen, M. Beyermann, M. Matsumoto, M. Hoshi, E. Hildebrand, and U. B. Kaupp, EMBO J. **24**, 2741 (2005).

[108] J. J. L. Higdon, J. Fluid Mech. **94**, 331 (1979).

[109] J. Elgeti, R. G. Winkler, and G. Gompper, SoftComp Newsletter No. 12 (2015), URL `http://www.eu-softcomp.net`.

[110] J. Elgeti and G. Gompper, EPL **101**, 48003 (2013).

[111] J. Elgeti and G. Gompper, EPL **85**, 38002 (2009).

[112] G. Li and J. X. Tang, Phys. Rev. Lett. **103**, 078101 (2009).

[113] A. C. Maggs, D. A. Huse, and S. Leibler, Europhys. Lett. **8**, 615 (1989).

[114] T. W. Burkhardt, J. Stat. Mech. p. P07004 (2007).

[115] H. H. Wensink and H. Löwen, Phys. Rev. E **78**, 031409 (2008).

[116] A. Wysocki, R. G. Winkler, and G. Gompper, EPL **105**, 48004 (2014).

[117] J. Stenhammar, D. Marenduzzo, R. J. Allen, and M. E. Cates, Soft Matter **10**, 1489 (2014).

[118] R. Ni, M. A. C. Stuart, and M. Dijkstra, Nat. Commun. **4**, 2704 (2013).

[119] F. Peruani, A. Deutsch, and M. Bär, Phys. Rev. E **74**, 030904(R) (2006).

[120] Y. Yang, V. Marceau, and G. Gompper, Phys. Rev. E **82**, 031904 (2010).

[121] F. Peruani, L. Schimansky-Geier, and M. Bär, Eur. Phys. J. Special Topics **191**, 173 (2010).

[122] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Löwen, and J. M. Yeomans, Proc. Natl. Acad. Sci. USA **109**, 14308 (2012).

[123] M. Abkenar, K. Marx, T. Auth, and G. Gompper, Phys. Rev. E **88**, 062314 (2013).

[124] A. Costanzo, J. Elgeti, T. Auth, G. Gompper, and M. Ripoll, EPL **107**, 36003 (2014).

[125] F. Hayashi, J. Insect Physiol. **42**, 859 (1996).

[126] H. D. M. Moore and D. A. Taggart, Biol. Reprod. **52**, 947 (1995).

[127] F. Hayashi, Funct. Ecol. **12**, 347 (1998).

[128] S. Immler, H. D. M. Moore, W. G. Breed, and T. R. Birkhead, PLoS ONE **2**, e170 (2007).

[129] H. D. M. Moore, K. Dvoráková, N. Jenkins, and W. G. Breed, Nature **418**, 174 (2002).

[130] I. H. Riedel, K. Kruse, and J. Howard, Science **309**, 300 (2005).

[131] Y. Yang, F. Qiu, and G. Gompper, Phys. Rev. E **89**, 012720 (2014).

[132] H. Machemer, J. Exp. Biol. **57**, 239 (1972).

[133] M. Sleigh, *2.The Biology of Cilia and Flagella* (Pergamon Press, Oxford, 1962), chap. The Structure of Cilia, pp. 11–75.

[134] R. Rikmenspoel, Biophys. J. **16**, 445 (1976).

[135] D. R. Brumley, M. Polin, T. J. Pedley, and R. E. Goldstein, Phys. Rev. Lett. **109**, 268102 (2012).

[136] M. J. Kim, J. C. Bird, A. J. Van Parys, K. S. Breuer, and T. R. Powers, Proc. Natl. Acad. Sci. USA **100**, 15481 (2003).

[137] M. Kim and T. R. Powers, Phys. Rev. E **69**, 061910 (2004).

[138] M. Reichert and H. Stark, Eur. Phys. J. E **17**, 493 (2005).

[139] B. Qian, H. Jiang, D. A. Gagnon, K. S. Breuer, and T. R. Powers, Phys. Rev. E **80**, 061919 (2009).

[140] S. Y. Reigh, R. G. Winkler, and G. Gompper, Soft Matter **8**, 4363 (2012).

[141] S. Y. Reigh, R. G. Winkler, and G. Gompper, PLoS ONE **8**, e70868 (2013).

[142] M. C. Lagomarsino, P. Jona, and B. Bassetti, Phys. Rev. E **68**, 021908 (2003).

[143] A. Vilfan and F. Jülicher, Phys. Rev. Lett. **96**, 058102 (2006).

[144] P. Lenz and A. Ryskin, Phys. Biol. **3**, 285 (2006).

[145] T. Niedermayer, B. Eckhardt, and P. Lenz, Chaos **18**, 037128 (2008).

[146] N. Uchida and R. Golestanian, Phys. Rev. Lett. **104**, 178103 (2010).

[147] C. Wollin and H. Stark, Eur. Phys. J. E **34**, 42 (2011).

[148] S. Gueron, K. Levit-Gurevich, N. Liron, and J. J. Blum, Proc. Natl. Acad. Sci. USA **94**, 6001 (1997).

[149] S. Gueron and K. Levit-Gurevich, Proc. Nat. Acad. Sci. USA **96**, 12240 (1999).

[150] B. Guirao and J. Joanny, Biophys. J. **92**, 1900 (2007).

[151] Y. Yang, J. Elgeti, and G. Gompper, Phys. Rev. E **78**, 061903 (2008).

[152] J. Elgeti and G. Gompper, Proc. Natl. Acad. Sci. USA **110**, 4470 (2013).

[153] N. Osterman and A. Vilfan, Proc. Natl. Acad. Sci. USA **108**, 15727 (2011).

[154] C. Eloy and E. Lauga, Phys. Rev. Lett. **109**, 038101 (2012).

[155] S. Gueron and K. Levit-Gurevich, Proc. Biol. Sci **268**, 599 (2001).

[156] J. R. Blake, Proc. Camb. Philos. Soc. **70**, 303 (1971).

F 8 Mechanical properties of biological protein polymers

G.H. Koenderink
Systems Biophysics Department
FOM Institute AMOLF
Amsterdam, Netherlands
www.amolf.nl/gijsjekoenderink

Contents

1 Introduction **2**

2 Structure and functions of biological protein polymers **3**
2.1 Cytoskeletal polymers 3
2.2 Extracellular matrix polymers 5

3 Semiflexible polymer models **7**
3.1 Single polymer mechanics 7
3.2 Mechanics of entangled polymer networks 9
3.3 Mechanics of crosslinked polymer networks 10

4 Summary and outlook **11**

References **13**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

The cells and tissues in our body owe their shape and mechanical strength to internal frameworks of protein biopolymers. Cells are supported by a composite network of *cytoskeletal filaments*, while tissues are supported by an *extracellular matrix* that is composed of fibrous proteins such as collagen, elastin and fibronectin (see Figure 1). The cytoskeleton not only enables cells to resist mechanical forces, but it also actively generates forces by means of active filament (de)polymerization and the action of motor proteins. As a result, cells can autonomously adapt their shape and mechanical behavior. The cytoskeleton is a paradigmatic example of a growing class of soft condensed matter known as *active matter* [1]. The cytoskeleton inside the cells is physically anchored to the extracellular matrix by means of transmembrane proteins known as integrins. The extracellular matrix provides anchoring support to cells and helps to organize them into a tissue. The integrin-mediated connections enable cells to influence the spatial organization of the matrix by the application of contractile forces, while at the same time allowing them to actively probe the matrix rigidity (mechanosensing). Mechanical measurements on whole cells and tissues have demonstrated intriguing material properties, including strain-stiffening, a superior mechanical strength, and non-equilibrium behavior that originates from active stress generation by cells [2]. The nonlinear (strain-stiffening) response to external forces is thought to protect cells and tissues from mechanical damage.

To reveal the physical basis of these material properties, there has been extensive research on reconstituted model systems made of purified proteins [3]. The obvious advantage of such a bottom-up approach is that the molecular and structural complexity can be precisely controlled, making it easier to perform quantitative experiments and also to compare experimental results directly with predictions of theoretical or computational models. Experimental and theoretical studies of such simplified single-component biopolymer networks have shown that the nonlinear elasticity and superior strength of cells and tissues stems from the physical properties of these polymers. A common feature of cytoskeletal and extracellular matrix polymers is their large size (10-100 nm diameters and micron lengths) compared to standard synthetic polymers, and their high bending rigidity. These features ensure that they form space-filling elastic networks already at low volume fractions and that they exhibit strain-stiffening behavior [4]. This strain-stiffening response contrasts with the rather linear response of conventional synthetic polymers such as polyacrylamide. However, recent work demonstrated that supramolecular synthetic polymers can be designed to mimic the remarkable nonlinear rheology of cytoskeletal networks [5].

Most existing theoretical models of biopolymer networks treat the filaments as simple elastic beams or as semiflexible polymers, ignoring the internal molecular packing structure. However, there is experimental evidence for several polymers, including fibrin [6] and intermediate filaments [7], that the underlying molecular structure plays an essential part in the network mechanics, being essential for the polymer's resilience and extensibility. In this chapter, I will provide an overview of the mechanical properties of the protein polymers that are primarily responsible for cell and tissue mechanics. I will focus on passive mechanical properties; reviews on the active mechanical properties of cells and tissues can be found elsewhere [1]. This chapter is partly adapted from an earlier, more extensive, review [8].

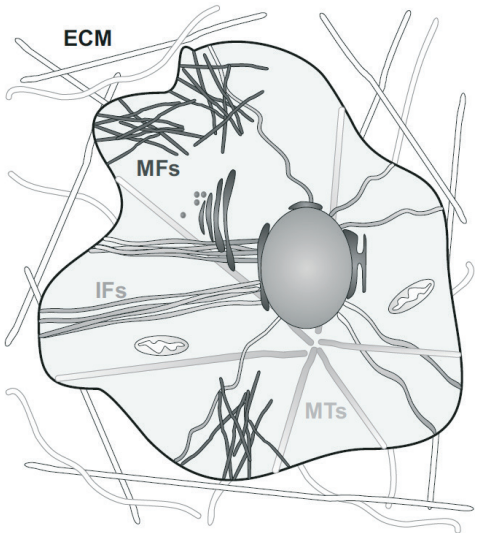


Fig. 1: *Schematic of biological polymers shaping the extracellular matrix (ECM) of animal tissues and those shaping cells (MTs, microtubules; IFs, intermediate filaments; MFs, actin microfilaments). The figure was prepared by I. Piechocka and is reprinted from [9].*

2 Structure and functions of biological protein polymers

2.1 Cytoskeletal polymers

Cytoskeletal polymers are supramolecular polymers built up from protein subunits linked via weak, non-covalent interactions. The specificity of these interactions results in highly ordered structures, while their non-covalent nature allows cytoskeletal protein polymers to (dis)assemble dynamically in response to biochemical or mechanical signals. This adaptability is crucial for processes such as cell migration, cell division, and tissue morphogenesis. A large number of accessory proteins such as molecular motors and crosslinkers organize the cytoskeletal polymers into higher-order structures that are tailored for specific tasks.

The cytoskeleton comprises three types of polymers that differ in structure, mechanical properties, and polymerization dynamics: actin filaments and microtubules, which are built of globular subunits, and intermediate filaments, which are built of fibrous subunits (Figure 2). All three polymers have lengths on the order of several μm and diameters of several nm (~10 nm for intermediate filaments, ~6 nm for actin filaments, and ~25 nm for microtubules). Actin and microtubules are conserved across the eukaryotic kingdom, being present in yeast, plant, and animal cells, whereas intermediate filaments are only present in animal cells.

Actin filaments are double-helical filaments formed by self-assembly of globular actin monomers (G-actin). The monomers are comprised of two domains separated by a cleft that binds a divalent cation and either adenosine triphosphate (ATP) or adenosine diphosphate

(ADP) [10]. The monomers assemble head-to-tail to form linear filaments that are structurally polar since the ligand-binding clefts of the monomers are all directed towards one end (denoted the „minus end“, while the other end is called the „plus end“). Actin monomers also associate via side-by-side contacts, forming a double-stranded helical structure with a 37-nm pitch [11]. The structural polarity of actin filaments is recognized by molecular motors (myosins), which take advantage of the polarity to move in a directional manner to transport cargo or slide (contract) actin filaments. Moreover, hydrolysis of the ATP bound to G-actin monomers that add onto the plus end of a growing filament provides chemical energy that maintains different monomer on- and off-rates at the two filament ends. The plus end has a higher on-rate than the minus end, leading to a phenomenon called treadmilling that allows actin filaments to exert polymerization forces to drive cell migration [12].

Accessory proteins organize actin filaments into different structures that can be classified as isotropic meshworks, branched networks, or stiff bundles [13]. In many cases, these structures are contractile due to the activity of myosin motor proteins. Right underneath the plasma membrane, actin and myosin for instance form a thin (100 nm) isotropic network known as the cortex, which is actively contracted by myosin motors [14]. During cytokinesis, the cortex transiently forms a contractile ring, which constricts and divides the cell. When adhered to a substrate, cells can actively migrate using a polymerization-based mechanism [15]. At the front of the cell, a thin, nearly two-dimensional branched array of actin filaments called the lamellipodium pushes the cell membrane forward, while at the back a contractile actin-myosin network exerts retraction forces to detach the cell from the substrate. In strongly adherent cells, actin and myosin form contractile bundles known as stress fibers, which span the cell and connect to focal adhesions [16]. In dense tissues, some cells can switch to alternative migration mechanisms involving myosin-induced membrane blebbing or matrix proteolysis [17, 18].

Like actin filaments, *microtubules* are made up of globular subunits, specifically heterodimers of α- and β-tubulin subunits. Both subunits bind guanine triphosphate (GTP) or guanine diphosphate (GDP), but the nucleotide binding site of the α-tubulin proteins is buried at the dimer interface [19]. The α-β-tubulin dimers assemble head-to-tail to form linear protofilaments, with α-tubulin at the “plus end” and β-tubulin at the “minus end”. Typically thirteen protofilaments associate side-by-side to form a hollow, cylindrical microtubule. Like actin filaments, microtubules are polar as well as non-equilibrium polymers, as a result of the hydrolysis of GTP bound to tubulin subunits that add onto the plus end of growing filaments. This leads to a phenomenon called dynamic instability, which allows microtubules to exert pushing and pulling forces, for instance at the cell cortex.

Due to their tubular structure and larger diameter, microtubules are stiffer than actin filaments by a factor of approximately 300 [20]. Their high rigidity allows microtubules to play a crucial role in organizing the cell interior and determining cell polarity. In interphase cells, microtubules usually emanate from a microtubule-organizing center positioned near the nucleus and grow radially outward toward the cell membrane, forming tracks for long-range cargo transport by kinesin and dynein motors [21]. When cells divide, the microtubules reorganize to form the mitotic spindle, which separates the chromosomes to the two daughter cells [22].

Intermediate filaments have a rather different structure from actin filaments and microtubules. The subunits are rod-shaped double-stranded parallel dimers with a length of 40-50 nm and a diameter of 2 nm. Intermediate filaments are encoded by 70 genes in the

human genome, which are often categorized into three assembly groups (I, II, III) that can co-exist as three separate systems within the same cell [23]. One type of intermediate filaments, the *lamins*, provides the membrane enveloping the nucleus with mechanical strength [24]. They polymerize by an assembly process involving head-to-tail polymerization [25]. The other intermediate filaments are found in the cytoplasm, and are expressed in a tissue-dependent and developmentally regulated manner. The most ubiquitously expressed intermediate filament types are *keratins* (in epithelial cells such as skin cells) and *vimentin* (in mesenchymal cells such as fibroblasts). Cytoplasmic IF proteins assemble into filaments via a multi-step pathway, where dimers first associate to form anti-parallel, approximately half-staggered tetramers, which laterally aggregate into unit-length filaments, which in turn longitudinally anneal to form filaments [23]. The fibrils subsequently mature by a compaction process. In contrast to actin filaments and microtubules, intermediate filaments are nonpolar: the two ends of the filament are identical. It is therefore believed that intermediate filaments do not serve as tracks for molecular motors. Instead, they are thought to endow cells with tissue-type dependent mechanical properties and with mechanical strength against large deformations.

2.2 Extracellular matrix polymers

The extracellular matrix is a composite meshwork of proteins and proteoglycans that forms the scaffold to which cells are adhered. There are two broad categories of extracellular matrices: basement membranes and connective tissue. Basement membranes are flat structures that provide a surface for attachment of polarized cell types such as epithelial and endothelial cells. In contrast, connective tissues provide a three-dimensional scaffold for cells. The main structural component of connective tissues is *collagen*. The collagen family encompasses 29 genetically distinct members, but the predominant collagen in all tissues except cartilage is type I collagen [26]. Collagen molecules consist of a remarkably long (300 nm) triple helix of three polypeptides, flanked by non-helical telopeptides (Fig. 2). Cells secrete collagen molecules with protective propeptides, which are removed in the extracellular space by enzymatic cleavage. This process triggers molecular self-assembly into stiff fibrils. The diameter and higher-order organization of the collagen fibers are tailored by auxiliary extracellular matrix components such as minor fibrillary collagens and small leucine rich proteoglycans to the specific functions of each tissue. In stiff tissues like tendon, collagen fibrils are aligned and have diameters around 200 nm to ensure tensile strength, whereas in the cornea, collagen fibrils form orthogonal meshworks and are only 30 nm thick to ensure optical transparency. It should be noted that, in addition to collagen, tissues contain a multitude of other structural macromolecules, which are expressed in a tissue-specific manner.

The molecular packing structure of collagen fibrils is remarkably ordered, with a precise axial stagger of 67 nm referred to as the “D-period”. The radial packing structure is more disordered, involving a densely interconnected assembly of right-handed, supertwisted pentamer microfibrils [27]. The axial packing order of the fibrils is essential to encode specific binding sites for cell adhesion receptors and extracellular matrix molecules. Moreover, the ordered structure is critical for the mechanical performance of tissues because it ensures that intermolecular crosslinks formed by lysyl oxidase - which are responsible for

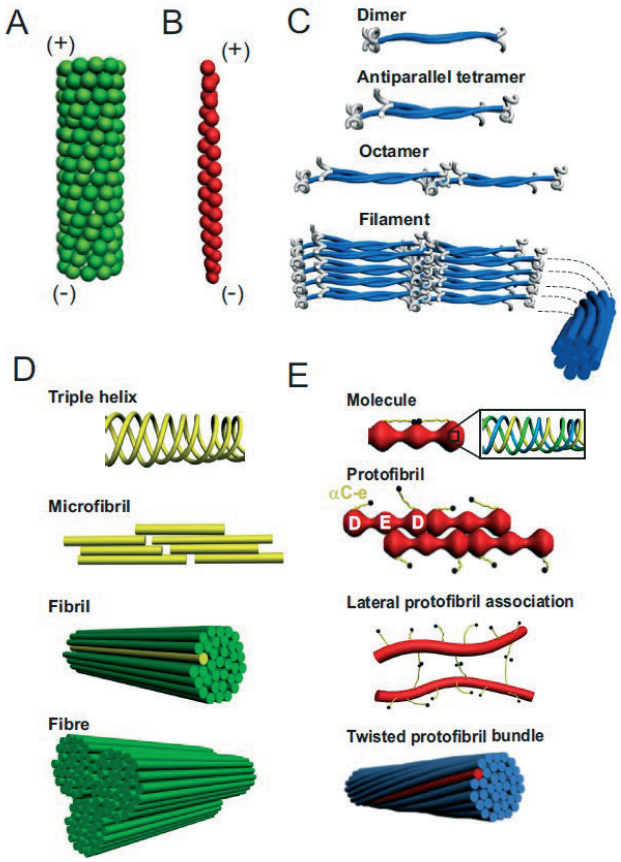


Fig. 2: *Supramolecular architecture of biological protein filaments. Microtubules (A) and actin filaments (B) are polar filaments built up from globular protein subunits. Intermediate filaments (C) are nonpolar due to an antiparallel arrangement of their fibrous subunits. Collagen fibres (D) are made of triple helical tropocollagen molecules (300 nm long and 1.5 nm wide) packed in an ordered, quarter-staggered arrangement. Fibrin fibers (E) are twisted bundles of loosely coupled protofibrils, which are themselves made of fibrous protein subunits in a half-staggered arrangement. The figure was prepared by I. Piechocka and is reprinted from [9].*

fibril integrity - are precisely localized to end-on molecular overlaps. This results in a specific crosslink spacing that provides an optimal balance between stiffness and toughness [28]. The hierarchical architecture of collagen fibers has made it difficult to dissect the individual contribution of each assembly level to the mechanical function. *In situ* X-ray scattering studies on tendons that are mechanically stretched suggest that all hierarchical levels contribute to the macroscopic response [29, 30].

A second protein filament in the extracellular matrix with very interesting mechanical properties is *fibrin*. Fibrin is a major protein in blood plasma that assembles into tough fibrous blood clots upon vascular injury. The immediate function of these clots is to stop blood loss, but they also serve as a temporary extracellular matrix that recruits cells that repair the wound [31]. Fibrinogen, the soluble precursor of fibrin, is a hexamer comprised of two sets of three

polypeptide chains, referred to as Aα, Bβ and γ (Fig. 2). The carboxy-terminal portion of each Aα-chain forms a compact αC-domain that is connected by an unstructured and flexible αC-connector region to the central region of the fibrinogen molecule. Polymerization is initiated by the enzyme thrombin, which cleaves two protective fibrinopeptides from the central E-domain, thus exposing so-called A- and B-knobs. The activated fibrin monomers spontaneously assemble in a half-staggered manner to form double-stranded protofibrils. This is encoded in non-covalent interactions of the A- and B-knobs with complementary a- and b-holes on the ends (distal D-domains) of adjacent fibrin molecules. The protofibrils subsequently bundle to form fibers. Bundling is mediated in part by interactions of the long and flexible αC-regions that project out from the surface of adjacent protofibrils. Fibrin fibers have a more open structure compared to collagen fibers, which gives them a lower bending rigidity. Biophysical studies have shown that fibrin networks are among the most resilient naturally-occurring polymer networks: fibrin networks can stiffen up to 1000-fold when sheared or stretched, and the individual fibers that build the networks can be stretched up to four-fold their original length before they break [6, 32].

3 Semiflexible polymer models

3.1 Single polymer mechanics

Mechanical models of cytoskeletal and extracellular matrix polymers usually ignore the molecular packing structure of the filaments, coarse-graining the filaments as (athermal) elastic beams or as semiflexible polymers. Semiflexible polymer models approximate the filaments by a smooth linear contour that resists bending with a quantity, κ, called the bending modulus. In the absence of thermal fluctuations, linear polymers would assume a straight shape. At finite temperatures, however, thermal fluctuations will cause these polymers to bend. The rigidity of semiflexible polymers is often quantified by the persistence length l_p, which is defined as the decay length of angular correlations along the polymer contour. The persistence length is related to the bending modulus by $\kappa = k_B T l_p$, where k_B is Boltzmann's constant and T is temperature. The persistence length can be easily measured by observing the thermal bending undulations of protein polymers by time-lapse fluorescence microscopy, as demonstrated for actin filaments, microtubules, and intermediate filaments [20, 33].

Based on the persistence length l_p relative to the contour length L, we can distinguish between three classes of polymers. If the polymer backbone offers little resistance to bending ($l_p << L$), the polymer is known as a *flexible polymer*. Since thermal fluctuations dominate, the polymer is bent into a coil whose conformation is well described by a fractal contour [34]. In the opposite limit ($l_p >> L$), the polymer is sufficiently stiff to resist thermal fluctuations and can be modeled as a *rigid rod* [35]. Collagen filaments fall into this category as a result of their large (~100 nm) diameter and tight molecular packing structure. A third, and intermediate, regime occurs when $l_p \sim L$. In this regime, thermal fluctuations cannot be neglected, and the polymer has a mostly straight shape but with undulations. Polymers in this intermediate regime are called *semiflexible*. Cytoskeletal filaments fall in this category, having contour lengths of several μm, and relatively large persistence lengths that range from 0.5-1 μm for intermediate filaments, to ~10 μm for actin filaments and several mm for microtubules. The small persistence length of intermediate filaments is surprising, given that their diameter (10 nm) is intermediate between those of actin filaments (6 nm) and

microtubules (25 mm). This suggests that the internal molecular packing structure of intermediate filaments affects their mechanics. It has been suggested that sliding between the fibrous subunits may act to lower the bending rigidity [7]. Despite their large (~100 nm) diameter, fibrin fibers also behave as semiflexible polymers. This is a result of their open bundle-like structure, made up of protofibrils separated by large water-filled interstices [6].

Because semiflexible polymers bend in response to thermal forces, their response to an applied pulling force is entropic in origin. Pulling straightens out the thermally-induced bends and thereby causes a reduction of the conformational entropy of the polymer [36]. The amplitude of thermally induced bends in the polymer depends on wavelength, typically quantified through the wave vector $q = n\pi/L$, where $n = 1, 2, 3$, etc. In the presence of an external pulling force f, the amplitude u_q of bending mode q is given by

$$\left\langle \left| u_q \right|^2 \right\rangle = \frac{2k_B T}{L\left(\kappa q^4 + fq^2\right)} \tag{1}$$

Long-wavelength bends (low q) have the largest bending amplitudes, while short-wavelength bends (high q) decay quickly, as q^{-4}. Applying a pulling force f reduces the transverse bending amplitudes u_q, resulting in an effective restoring force in the linear response regime given by:

$$f \propto \frac{l_p \kappa}{L^4} x \tag{2}$$

where x denotes the displacement of the end-to-end-distance vector of the polymer contour from its equilibrium length. The effective spring constant of a semiflexible polymer is thus $l_p \kappa L^{-4}$. This expression was experimentally verified with optical tweezers for actin filaments as well as microtubules [37]. Once the excess length stored in thermal fluctuations is pulled out, the polymer strongly resists further stretching, leading to strong stiffening. However, at large enough pulling forces, the polymer backbone can be stretched. The entropic spring model can be generalized to this situation by introducing an enthalpic quantity called the stretch modulus, μ [4, 38]. Experimental evidence suggests that actin filaments and microtubules break before they are substantially extended, whereas intermediate filaments and fibrin fibers can be stretched up to 4-fold before they break [7, 32]. Fiber breakage and extensibility is governed by the internal packing structure of the fibers. In case of intermediate filaments and fibrin fibers, the bundle-like packing of the fibrous subunits apparently promotes extensibility by mechanisms involving subunit sliding and/or forced unfolding [7, 32, 39, 40].

The force-extension relation of semiflexible polymers is highly asymmetric. Under a compressive force, semiflexible polymers readily undergo a buckling instability once the load exceeds the critical Euler force, f_c [35]:

$$f_c \propto \kappa L^{-2} \tag{3}$$

Actin filaments can thus withstand pulling forces but not pushing forces. However, bundling as well as coupling to other cytoskeletal structures can reinforce the filaments against compressive loads [41, 42]. Importantly, the buckling force sets a limit on the maximal protrusive force that growing actin filaments or microtubules can exert. Given that microtubules are 300-fold stiffer than actin filaments, they are more stable against

compressive forces [43]. Some theoretical models of cell mechanics therefore assume that microtubules act as compressive struts whereas actin filaments and intermediate filaments act as tensile elements [44].

3.2 Mechanics of entangled polymer networks

The viscoelastic properties of a solution of semiflexible polymers are strongly dependent on the polymer concentration. In the dilute regime, when the polymer number density c is much less than L^{-3}, neighboring filaments are spaced far enough apart that they do not significantly interfere with each other's motions. However, once the number density increases beyond L^{-3}, the filaments start to interact via excluded volume interactions. In this entangled regime, the diffusion of each rod is constrained by its neighbors, as modeled by the classical reptation model of Doi and Edwards [45]. In this model, each filament is confined to an elongated virtual tube formed by the presence of neighboring rods and slides back and forth along this tube in a snake-like motion called reptation. The tube model was directly confirmed via experiments on actin filaments, by tracking the diffusion of fluorescently labeled filaments in a dense (0.1 to 2 mg mL^{-1}) background of unlabeled actin filaments [46].

Entangled polymer networks are viscoelastic, exhibiting a frequency-dependent rheological response with a viscous as well as an elastic component. The response of a viscoelastic material to an applied shear stress is given by the complex shear modulus, $G^* = G' + iG''$, where G' is the storage modulus measuring elastic behavior while G'' is the loss modulus measuring the viscous behavior. Experimentally, G^* is usually determined by rheometry, applying a small oscillatory shear strain of controlled frequency to the polymer solution contained between two concentric plates, and measuring the resultant oscillatory shear stress response. The stress/strain ratio provides G^*, while the phase shift between the stress and strain signals reflects the extent of viscous dissipation. At small frequencies, corresponding to times longer than the reptation time of the filaments, entangled polymer solutions behave as viscous liquids, whereas at higher frequencies, the solutions behave as elastic solids. For actin filaments, which have typical lengths of several micrometers, the reptation time scale is on the order of minutes to hours [47]. Theoretical models predict, based on scaling arguments, that the elastic modulus depends on two concentration-dependent length scales [48] according to:

$$G' \propto \frac{k_B T}{\xi^2 l_e} \tag{4}$$

The two length scales are the mesh size ξ, and the entanglement length l_e, which describes the typical length over which filament entanglements restrict thermal fluctuations. The mesh size scales as $\xi \sim \varphi^{-1/2}\, a$ (where a is the filament diameter) for random networks of rigid filaments [49], while the entanglement length scales as $l_e \sim (a^4\, l_p)^{1/5}\, \varphi^{-2/5}$ [50]. Thus, G' should increase with polymer volume fraction as $\varphi^{7/5}$. This prediction was experimentally verified for entangled actin networks [47]. Intermediate filaments and microtubules exhibit more complicated dependencies of G' on filament density as a result of attractive filament interactions that depend on temperature and added salts [51, 52].

3.3 Mechanics of crosslinked polymer networks

Both intracellular and extracellular protein filament networks are usually crosslinked. In case of extracellular polymers such as collagen and fibrin, these crosslinks are covalent molecular bonds created by enzymes. Moreover, fibrin and collagen networks are usually branched, exhibiting an average coordination number close to 3 [53, 54]. In case of the cytoskeleton, crosslinks are transient, because they are created by specialized crosslinking proteins that transiently and specifically bind to the cytoskeletal filaments. Examples of well-studied crosslinking molecules are the homodimeric actin crosslinkers filamin and α-actinin and the monomeric crosslinker fascin [13]. Experiments with purified actin have shown that these crosslinkers can organize actin filaments into bundles, isotropic networks, or mixed network/bundle phases, depending on crosslinker concentration and on the molecular structure and size of the crosslinker molecule. There have been efforts to theoretically predict the network structure by analytical approaches and computer simulations [55], but the interplay of crosslinker protein structure and binding kinetics is still not entirely clear. Theoretically predicting network structures is further hampered by the possibility of nonequilibrium network structures [56].

Introducing crosslinks in a network of entangled semiflexible polymers introduces a new length scale important for the network mechanics, called the crosslink distance l_c. This distance is a key parameter that determines whether a macroscopically imposed network deformation locally results in filament stretching or filament bending. When a macroscopic shear strain results predominantly in filament stretching, the network experiences affine (uniform) deformations. Since all filaments experience exactly the same deformations, the network elasticity can in this case be calculated analytically by orientationally averaging over the force-extension response of each filament [4]. Provided that thermal fluctuations are appreciable and the network is isotropic and randomly crosslinked, the storage modulus depends on the crosslink concentration (molar ratio of crosslinks to actin monomers) c_x according to [57]:

$$G'_{affine} \propto c_x \kappa l_p l_c^{-3} \tag{5}$$

This relation was experimentally verified for actin filaments crosslinked with the protein scruin [57] as well as for intermediate filaments crosslinked with divalent cations [58].

When the filaments are very stiff and/or when the network connectivity is low, it will be more energetically favorable for the filaments (or bundles) to bend rather than stretch when a macroscopic shear stress is applied, resulting in nonaffine deformations. Compared to an affinely deforming networks, the network stiffness will always be lower since nonaffinity increases the number of degrees of freedom in the system. In this case, it is much more challenging to theoretically predict the elastic modulus of the network. This regime has therefore mostly been explored by computer simulations of Mikado models (randomly deposited rigid rods) and disordered lattices. Since the connectivity of biological polymer networks is usually somewhere between 3 (for a branched networks) and 4 (for a crosslinked network), they are in principle in a sub-marginal state: their average connectivity is lower than the Maxwell criterion, which states that networks of springs are stable only when the average connectivity (in 3D) is at least 6. Computer simulations demonstrated that biopolymer networks are nevertheless stable, due to the high bending rigidity of the filaments [59]. Also stresses, such as tensile forces applied by motors, can stabilize initially floppy (submarginal) networks [60]. The simulations predict a strong dependence of the network

mechanics on fiber length. The bending-dominated (nonaffine) elastic regime is controlled by fiber length, and for long enough filaments there is a crossover to a stretch-dominated (affine) regime [61].

When biopolymer networks are subjected to large shear stresses, they tend to stiffen strongly. This nonlinear response poses technical challenges for measuring the rheological properties in a quantitative fashion. When a large oscillatory stress is applied, the strain response tends to deviate strongly from a sinusoidal shape and the shear modulus extracted from the stress-strain ratio represents only the first harmonic component of the response. Most studies on actin networks therefore report instead the differential storage modulus, $K' = [\delta\sigma / \delta\gamma] |\sigma_0$, which is the local tangent of the stress-strain curve. K'-values can be directly compared with theoretical predictions based on semiflexible polymer models [62, 63]. K' is usually obtained by a differential prestress (also known as parallel superposition) protocol, where small amplitude oscillations are superimposed on a steady-state shear flow. An alternative protocol, which is more suitable for materials exhibiting creep such as entangled actin solutions, is a strain rate ramp protocol, where K' is extracted by differentiating the stress/strain curve measured at different strain rates.

In the limit of semiflexible polymers, where thermal fluctuations are appreciable, the origin of the strain-stiffening behavior is the asymmetric, entropic force-extension behavior of the filaments, which are easily buckled but strongly resist stretching. Provided that the networks deform in an affine manner, the differential elastic modulus is predicted to increase with stress σ according to $K' \sim \sigma^{3/2}$ [57]. This prediction was indeed confirmed in measurements on actin networks crosslinked with scruin [57] and intermediate filament networks crosslinked by divalent cations [58]. In case of actin networks, there have been a few studies that showed how the strain-stiffening response is modified by molecular properties of the crosslinker protein such as its mechanical compliance [64, 65] and its force-dependent dissociation kinetics [66]. For networks of intermediate filaments and fibrin fibers, a more complex strain-stiffening response is observed, where the initial entropic stiffening regime is followed by an enthalpic regime that originates from the extensibility of the fiber backbone [6, 58]. In case of fibrin, the networks furthermore exhibit a distinct strain-stiffening regime at high strain, which reflects the fact that the polymers themselves stiffen when they are highly stretched. This fiber stiffening response is probably a result of force-induced monomer unfolding [32].

Networks of rigid actin bundles [67] or collagen fibers [68], for which thermal fluctuations are negligible, also tend to strain-stiffen. In this case, computer simulations indicate that strain-stiffening is a consequence of a transition from bending-dominated (nonaffine) elasticity at small strain to stretching-dominated (affine) elasticity at large strain [61, 69, 70]. Imaging of sheared networks by confocal fluorescence microscopy has indeed provided direct evidence of geometric realignment of these networks [39, 71, 72].

4 Summary and outlook

In this chapter, I have provided an introduction to the passive mechanical properties of cytoskeletal and extracellular matrix biopolymers based on experimental insights gained by studying simple purified model systems, theoretical modeling, and computer simulations. Actin filaments and microtubules, which are comprised of globular subunits held together by

weak lateral and longitudinal bonds, are generally well approximated as semiflexible polymers. Intermediate filaments and fibrin fibers, which are rope-like bundles of fibrous subunits, exhibit more complex mechanical properties. Like actin filaments and microtubules, they exhibit entropic elasticity for small strains. However, at large strains, they exhibit elastomeric behavior that somehow originates from their hierarchical molecular structure. Networks of stiff polymers like actin bundles or collagen fibers also exhibit highly non-linear mechanical properties, but in this case the network mechanics is governed by nonaffine bending at small strain and affine stretching at high strain.

Dissecting the contributions of each level of hierarchy to the macroscopic mechanical response of biopolymer networks poses major experimental as well as theoretical challenges, requiring multiscale approaches. On the experimental side, combinations of macroscopic mechanical measurements with *in situ* X-ray scattering [29, 30, 32] or with IR-vibrational spectroscopy [73] provide a promising route. Moreover, with recent technical advances in biophysics and nanotechnology such as optical tweezers, atomic force microscopy, and microelectromechanical devices, it is now possible to measure mechanical properties of individual molecular subunits and individual fibrils [37, 74, 75]. On the theoretical side, an interesting way forward to model a multiscale system such as actin bundles or fibrin fibers is to treat the filaments as semiflexible polymer bundles. This approach allows one to account for the details of the coupling between the constituent monomers [76] and for crosslink reversibility [77]. In addition, computer simulations will greatly benefit from systematic coarse-graining approaches in order to simulate multi-scale hierarchical polymers such as collagen from the atomic up to the fiber/network scale [78]. Another key challenge for the future is to bridge between the relative simplicity of one-component polymer networks and the complexity of whole tissues and cells. This involves considering the synergistic properties of composite networks of polymers of distinct bending rigidities [79] as well as the role of active driving forces coming from molecular motors and non-equilibrium polymerization dynamics [80]. Moreover, biophysicists are increasingly turning to quantitative experiments and coarse-grained modeling of entire living cells or even model organisms such as *C. Elegans* embryos [81, 82].

References

[1] M. Marchetti, J. Joanny, S. Ramaswamy, T. Liverpool, J. Prost, M. Rao, and R. Simha, Rev. Mod. Phys. 85, 1143 (2013)
[2] Y. Fung, Biomechanics: Mechanical properties of living tissues (Springer, 1993)
[3] A. Liu and D. Fletcher, Nat. Rev. Mol. Cell Biol. 10, 644 (2009)
[4] C. Storm, J. Pastore, F. MacKintosh, T. Lubensky, and P. Janmey, Nature 435, 191 (2005)
[5] M. Jaspers, M. Dennison, M. Mabesoone, F. MacKintosh, A. Rowan, and P. Kouwer, Nat. Commun. 5, 5808 (2014)
[6] I. Piechocka, R. Bacabac, M.M. Potters, FC, and G. Koenderink, Biophys. J. 98, 22281 (2010)
[7] L. Kreplak and D. Fudge, Bioessays 29, 26 (2007)
[8] J. Alvarado and G. Koenderink, in: Cell and Matrix Mechanics, Edited by R. Kaunas and A. Zemel (Taylor & Francis, London, 2014)
[9] I. Piechocka, Biopolymers: from structural hierarchy to nonlinear rheology (PhD thesis, VU University Amsterdam, 2011)
[10] M. Carlier, C. Valentin-Ranc, C. Combeau, S. Fievez, and D. Pantoloni, Adv. Exp. Med. Biol. 358, 71 (1994)
[11] V. Galkin, A. Orlova, G. Schröder, and E. Egelman, Nature Struct. Mol. Biol. 17, 1318 (2010)
[12] T. Pollard and G. Borisy, Cell 112, 453 (2003)
[13] O. Lieleg, M. Claessens, and A. Bausch, Soft Matter 6, 218 (2010)
[14] G. Salbreux, G. Charras, and E. Paluch, Trends Cell Biol. 22, 536 (2012)
[15] M. Abercrombie, Proc. Royal Soc. B: Biological Sciences 207, 129 (1980)
[16] S. Tojkander, G. Gateva, and P. Lappalainen, J. Cell Sci. 125, 1855 (2012)
[17] E. Paluch and E. Raz, Curr. Opin. Cell Biol. 25, 582 (2013)
[18] P. Friedl, E. Sahai, S. Weiss, and K. Yamada, Nat. Rev. Mol. Cell Biol. 13, 743 (2012)
[19] M. Gardner, M. Zanic, and J. Howard, Curr. Opin. Cell Biol. 25, 14 (2013)
[20] F. Gittes, B. Mickey, J. Nettleton, and J. Howard, J. Cell Biol. 120, 923 (1993)
[21] K. Barlan, M. Rossow, and V. Gelfand, Curr. Opin. Cell Biol. 25, 483 (2013)
[22] K. Helmke, R. Heald, and J. Wilbur, Int. Rev. Cell Mol. Biol. 306, 83 (2013)
[23] H. Herrmann, H. Bär, L. Kreplak, S. Strelkov, and U. Aebi, Nat. Rev. Mol. Cell Biol. 8, 562 (2007)
[24] Y. Gruenbaum and O. Medalia, Curr. Opin. Cell Biol. 32C, 7 (2014)
[25] L. Kapinos, J. Schumacher, N. Mücke, G. Machaidze, P. Burkhard, U. Aebi, S. Strelkov, and H. Herrmann, J. Mol. Biol. 396, 719 (2010)
[26] P. Fratzl, ed. Collagen. Structure and mechanics (Springer: New York.2008)
[27] J. Orgel, T. Irving, A. Miller, and T. Wess, Proc. Natl. Acad. Sci. USA 103, 9001 (2006)
[28] M. Buehler, Proc. Natl. Acad. Sci. USA 103, 12285 (2006)
[29] N. Sasaki and S. Odajima, J. Biomech. 29, 1131 (1996)
[30] K. Misof and G.F. Rapp, P, Biophys. J 72, 1376 (1997)
[31] P. Janmey, J. Winer, and J. Weisel, J. Royal Soc. Interface 6, 1 (2009)
[32] A. Brown, R. Litvinov, D. Discher, P. Purohit, and J. Weisel, Science 325, 741 (2009)
[33] B. Noding and S. Köster, Phys. Rev. Lett. 108, 088101 (2012)
[34] P. deGennes, Scaling Concepts in Polymer Physics (Cornell University Press, 1979)

[35] L. Landau, E. Lifshitz, A. Kosevich, and L. Pitaevskiĭ, Theory of Elasticity (Butterworth-Heinemann, 1986)
[36] F. MacKintosh, J. Käs, and P. Janmey, Phys. Rev. Lett. 75, 4425 (1995)
[37] J. vanMameren, K.G. Vermeulen, F, and C. Schmidt, J. Phys. Chem. B 113, 3837 (2009)
[38] T. Odijk, Macromolecules 28, 7016 (1995)
[39] S. Münster, L. Jawerth, B. Leslie, J. Weitz, B. Fabry, and D. Weitz, Proc. Natl. Acad. Sci. USA 110, 12197 (2013)
[40] R. Kirmse, Z. Qin, C. Weinert, A. Hoenger, M. Buehler, and L. Kreplak, PLoS One 5, e12115 (2010)
[41] M. Das and F. MacKintosh, Phys. Rev. Lett. 105, 138102 (2010)
[42] M. Footer, J. Kerssemakers, and J.D. Theriot, M, Proc. Natl. Acad. Sci. USA 104, 2181 (2007)
[43] M. Dogterom, J. Kerssemakers, G. Romet-Lemonne, and M. Janson, Curr. Opin. Cell Biol. 17, 67 (2005)
[44] D. Ingber, W. Ning, and S. Dimitrije, Rep. Progr. Phys. 77, 046603 (2014)
[45] M. Doi and S. Edwards, J. Chem Soc. Faraday Trans. 2 74, 1789 (1978)
[46] J. Käs, H. Strey, and E. Sackmann, Nature 368, 226 (1994)
[47] B. Hinner, M. Tempel, E. Sackmann, K. Kroy, and E. Frey, Phys. Rev. Lett. 81, 2614 (1998)
[48] C. Broedersz and F. MacKintosh, Rev. Mod. Phys. 86, 995 (2014)
[49] C.B. Schmidt, M, G. Isenberg, and E. Sackmann, Macromolecules 22, 3638 (1989)
[50] H. Isambert and A. Maggs, Macromolecules 29, 1036 (1996)
[51] P. Pawelzyk, N. Mücke, H. Herrmann, and N. Willenbacher, PLoS One 9, e93194 (2014)
[52] Y. Lin, G. Koenderink, F. MacKintosh, and D. Weitz, Macromolecules 40, 7714 (2007)
[53] T. Baradet, J. Haselgrove, and J. Weisel, Biophys. J 68, 1551 (1995)
[54] S. Lindström, D.K. Vader, A, and D. Weitz, Phys. Rev. E 82, 051905 (2010)
[55] I. Borukhov, R. Bruinsma, W. Gelbart, and A. Liu, Proc. Natl. Acad. Sci. USA 102, 3673 (2005)
[56] T. Falzone, M. Lenz, D. Kovar, and M. Gardel, Nature Comm. 3, 861 (2012)
[57] M. Gardel, J. Shin, F. MacKintosh, L. Mahadevan, P. Matsudaira, and D. Weitz, Science 304, 1301 (2004)
[58] Y. Lin, N. Yao, C. Broedersz, H. Herrmann, F. Mackintosh, and D. Weitz, Phys. Rev. Lett. 104, 058101 (2010)
[59] C. Broedersz, X. Mao, T. Lubensky, and F. MacKintosh, Nature Phys, 7, 983 (2011)
[60] M. Sheinman, C. Broedersz, and F. MacKintosh, Phys. Rev. Lett. 109, 238101 (2012)
[61] C. Broedersz, M. Sheinman, and F. MacKintosh, Phys. Rev. Lett. 108, 078102 (2012)
[62] C. Semmrich, R. Larsen, and A. Bausch, Soft Matter 4, 1675 (2008)
[63] C. Broedersz, K. Kasza, L. Jawert, S. Münster, D. Weitz, and F. MacKintosh, Soft Matter 6, 4120 (2010)
[64] M. Gardel, F. Nakamura, J. Hartwig, J. Crocker, T. Stossel, and D. Weitz, Proc. Natl. Acad. Sci. USA 103, 1762 (2006)
[65] A. Sharma, M.H. Sheinman, KM, and F. MacKintosh, Phys. Rev. E 88, 052705 (2013)
[66] N.B. Yao, CP, M.B. Depken, DJ, M. Pollak, F. Mackintosh, and D. Weitz, Phys. Rev. Lett. 110, 018103 (2013)
[67] O. Lieleg, M. Claessens, C. Heussinger, E. Frey, and A. Bausch, Phys. Rev. Lett. 99, 088102 (2007)

[68] A. Stein, D. Vader, D. Weitz, and L. Sander, Complexity 16, 22 (2011)
[69] P. Onck, T. Koeman, T. vanDillen, and E. vanderGiessen, Phys. Rev. Lett. 95, 178102 (2005)
[70] E. Conti and F. Mackintosh, Phys. Rev. Lett. 102, 088102 (2009)
[71] K. Schmoller, P. Fernandez, R. Arevalo, D. Blair, and A. Bausch, Nat. Commun. 1, 134 (2010)
[72] D. Vader, A. Kabla, D. Weitz, and L. Mahadevan, PLoS One 4, e5902 (2009)
[73] R. Litvinov, D. Faizullin, Y. Zuev, and J. Weisel, Biophys. J 103, 1020 (2012)
[74] L. Yang, K. vanderWerf, C. Fitié, M. Bennink, P. Dijkstra, and J. Feijen, Biophys. J 94, 2204 (2008)
[75] Z. Shen, M. Dodge, H. Kahn, R. Ballarini, and S. Eppell, Biophys. J. 99, 1986 (2010)
[76] C. Heussinger, M. Bathe, and E. Frey, Phys. Rev. Lett. 99, 048101 (2007)
[77] R. Vink and C. Heussinger, J. Chem. Phys. 136, 035102 (2012)
[78] A. Gautieri, S. Vesentini, A. Redaelli, and M. Buehler, Nano Lett. 11, 757 (2011)
[79] F. Huber, A. Boire, M. López, and G. Koenderink, Curr. Opin. Cell Biol. 32C, 39 (2014)
[80] F. Mackintosh and C. Schmidt, Curr. Opin. Cell Biol. 22, 29 (2010)
[81] M. Mayer, M. Depken, J. Bois, F. Jülicher, and S. Grill, Nature 467, 617 (2010)
[82] J. Sedzinski, M. Biro, A. Oswald, J. Tinevez, G. Salbreux, and E. Paluch, Nature 476, 462 (2011)

G

FUTURE TECHNOLOGIES

G 1 Self-Healing Polymers

W. Pyckhout-Hintzen
Jülich Centre for Neutron Science JCNS-1
Forschungszentrum Jülich GmbH

Contents

1 Introduction 2

2 Dynamic Bonds 3

2.1 Non-Dynamic Covalent Bonding-Based Self-Healing 4
2.2 Dynamic Covalent Bonding-Based Self-Healing 4
2.3 Dynamic Non-Covalent Bonding-Based Self-Healing 6

3 Physical Characterization and Efficiency 10

4 Future Directions 13

References 14

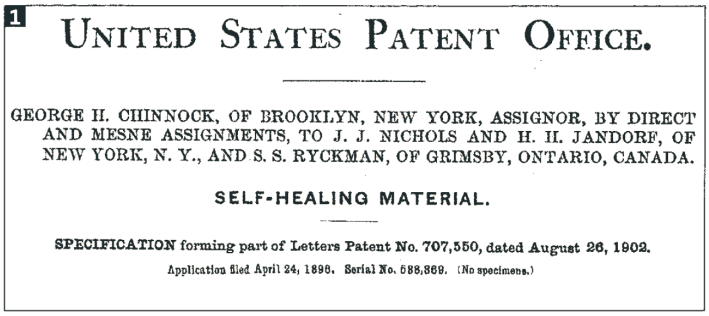
Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

As a consequence of evolution biological materials like human skin are the most advanced functional systems. They differ from synthetic materials that we know in everyday life at least in their ability to considerable self-repair [1,2,6,10,11]. Natural materials thus continuously adapt and so are responsive to damage. It would not be the first time that mankind is making use of this wealth of examples in nature e.g. to construct buildings or develop new machinery. Rubbers are used in many everyday applications as sealing elements or joints, vibration damper and many more. During use they are subjected to fatigue, impact, abrasion, wear etc., which causes formation of small fractures. Those can lead to irreparable failure, limiting the life time and requireing frequent replacement. This causes significant investments and waste of time. Including self-repair concepts to elastomers would be beneficial to prolongate their service time. This is especially important in areas where repair or replacement of parts is very costly and impractical e.g. aeronautic and astronautic applications. Unexpected damages from impacts of dust and interstellar rocks in space require strong innovative materials that ideally can repair itself. Hairline cracks cause major damage over time and lead to cata-strophic failure [1]. In this time-limited short survey we aim at scratching the rich field of current self-repairing matters and methods. Furthermore, this review will be entirely restricted to polymeric materials, although there are many more applications. We leave out concrete, ceramics and even nanocomposites [11]. The idea of self-repair is not at all new and attempts can be traced back to the early days of synthetic polymer chemistry more than a century ago, especially in the field of tire development. Fig 1 shows the heading of a patent specification, that was filed in the year 1896. Therefore, in the present review, the most common self-healing strategies, their classification and properties will be covered without having the claim to be complete. We strongly suggest the interested reader to look up the cited, exciting and for sure inspiring literature and references therein to get a complete picture of the broad field of self-healing materials [5,6,8,11].
Designing novel materials for the future is an inter-disciplinary process and requires the joint efforts of microscopic and macroscopic studies into the fundamental principles. With some intelligence one may arrive at a knowledge-based transfer between the different levels of complexity that distinguish nature and well-defined molecular details.

As we will discuss in the manuscript, the possibility and the concept of self-healing is associated with at least a partial mobility in a polymeric matrix. A molecular ‘glue’ is required which rejoins the fractured sections of cracks. This molecular glue always needs a combination of physical and chemical healing principles. Ideally, damage sensors or triggers would detect the damage automatically and could initiate the healing process, either by transport of the healing agent into the troubled zone or by entering into the mending reaction itself. Skin e.g contains millions of microscopic nerves and nerve endings that signal the brain. Often this self-monitoring allows one to react on time to the warning and to avoid serious injury. On the long-term, nanometer-scaled sensors in materials to self-monitor would be highly beneficial. Damage *prevention* is certainly good but the question is whether the amplitude of damage or its impact can be predicted in the necessary accuracy? Therefore, a more important property is damage *management*, which relates to the intrinsic capability to repair the generated damage *in situ*. Ideally, microcracks are efficiently repaired and the original function is restored quantitatively to grant the “life-long” performance observed in skin tissue.

The manuscript is divided in following sections: first, we will discuss self-healing in polymers along a concept called dynamic bonding, in various aspects. Thereby we will refrain from many chemical details and refer e.g. to the contribution about *Functional Polymers* by M. Möller (RWTH Aachen). Instead, some generic examples will be strongly simplified and highlighted. After this, attention is given to the physical characterization or validation of the mechanisms and finally we switch to some current applications and show up future directions. During the lecture the matter will be illustrated by some impressive video material and on-going research at the JCNS-1.



1

UNITED STATES PATENT OFFICE.

GEORGE H. CHINNOCK, OF BROOKLYN, NEW YORK, ASSIGNOR, BY DIRECT AND MESNE ASSIGNMENTS, TO J. J. NICHOLS AND H. H. JANDORF, OF NEW YORK, N. Y., AND S. S. RYCKMAN, OF GRIMSBY, ONTARIO, CANADA.

SELF-HEALING MATERIAL.

SPECIFICATION forming part of Letters Patent No. 707,550, dated August 26, 1902.

Application filed April 24, 1896. Serial No. 588,869. (No specimens.)

Fig. 1: *Filed patent in self-healing dating back to 1896*

2 Dynamic Bonds

Before we diversify we note that there are several ways to classify self-healing processes in polymers [2,7,8,11]. For the sake of generality we will not distinguish them according to monomeric type nor application. The ability of a material to heal damages leads to mainly two classes: **autonomously** repairing polymers, without any external mostly human intervention or stimulus, and **non-autonomously** healing after an impulse is given. Sometimes also the terms **"extrinsic"** and **"intrinsic"** self-healing concepts are found. In case of extrinsic self-healing materials, the healing process is based on external healing components (i.e. micro- or nanocapsules) intentionally pre-embedded into the matrix materials which itself is inactive. These capsules can provide the mobile phase as a result of damage. Alternatively, intrinsic self-healing requires no separate healing agent. The molecular mechanisms involved in this healing proces consist of physical interactions and chemical interactions. The system utilizes an inherent material triggered either by a damage event or in combination with an external stimulus. The classification is not well defined and we have opted here to argue along the concept of dynamic bonding which to our opinion describes in the best way the mechanisms of self-repair via local chemistry. This precludes the already mentioned most famous and old mechanism of autonomous extrinsic repair. However, to maintain the review character of this contribution, it will be shortly reviewed as well.

Any type of bond that undergoes reversible breaking and reformation selectively, under equilibrium conditions, can be termed a "dynamic bond" [2]. It can be either self-complementary i.e. linking the same functionality, or hetero-complementary if two different functional groups are concerned. We distinguish in the following "dynamic covalent bonding" and "dynamic non-covalent bonding" and refer to the other as "non-dynamic covalent bonding". Both concepts of dynamic and non-dynamic covalent bonding act on different

length scales: whereas chemical bonding is a typical molecular scale approach and depends on the local chemistry of the partners, the encapsulation of solvents, healing agents, monomer and catalysts is mesoscopic in nature and deals with at least micron-sized bubbles or channels.

2.1 Non-Dynamic Covalent Bonding-Based Self-Healing

The encapsulation approach has been mainly developed for crosslinked materials [7]. The underlying mechanism is called **autonomous** since the repair is triggered by the damaging process itself. However, it is a good example of an irreversible covalent bond formation. Fig. 2 illustrates shortly the process: a microcapsule is ripped open by a propagating crack after which the contents are either set free into the matrix where the reaction can proceed locally, or a catalyst is drawn into the microcapsule by capillary forces and crosslinks the interior of the capsule. Cracks can be effectively stopped in this way. A major disadvantage is that the healing process is one-time-only and thus cannot be repeated. Furthermore, attention should be paid not to pre-damage the capsules during polymer processing. One classical system uses a catalytic metathesis reaction with a ruthenium-based Grubbs catalyst and dicyclopentadiene monomers inside a 10-10000µm-large microcapsule. The ring-opening methathesis polymerization (ROMP)proceeds as long as 30°C<T<120°C. This rather ambient temperature window strongly limits new applications in more harsh conditions and environments.

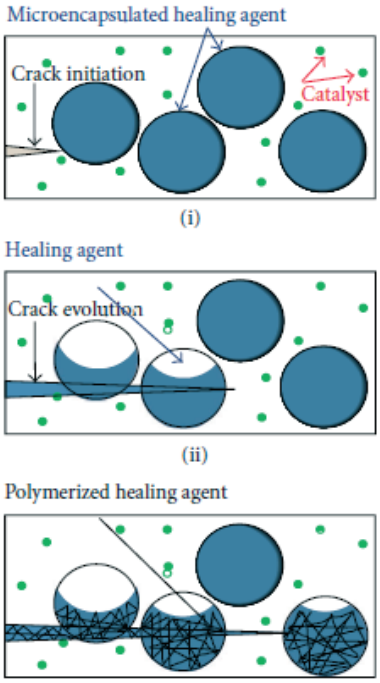


Fig. 2: *Encapsulation: a crack is stopped by post-polymerization [11]*

2.2 Dynamic Covalent Bonding-Based Self-Healing

One approach to introduce processability into cross-linked polymers is the introduction of reversible cross-links in polymeric systems [8,10,11]. The nature of dynamic covalent cross-links, whether they are at equilibrium or triggered via an external stimulus, is an important

issue if a material is to be tailored for a specific application. In addition to re-fabrication and recyclability, reversible cross-links exhibit self-healing properties. However, a reversible cross-linked system does not show self-repairing ability by its own. An external trigger such as thermal, photo-, or chemical activation is needed to achieve reversibility, and thereby the self-healing ability. Thus, these systems show a **non-autonomic** healing phenomenon. Such dynamic covalent bonds then behave more or less like traditional covalent bonds. The material in which they are embedded, however, can still exploit the inherent dynamic features.

- **Diels-Alder**:

The Diels-Alder cycloaddition is of interest for repair mechanisms in polymers due to the existence of its reverse or retro-reaction. Whereas the 4+2 cycloaddition of a 1,3-diene and a dienophilic group already proceeds at almost ambient conditions through a 6-center mechanism, leading to a coupled product (see Fig. 3) , increasing the temperature typically above 100°C reverses the path way. The self-healing activity is thus connected with the 2 groups which are either incorporated in side chains of the backbone polymer or make up the backbone itself [7]. Despite its success and ease of reaction the efficiency is rather low.

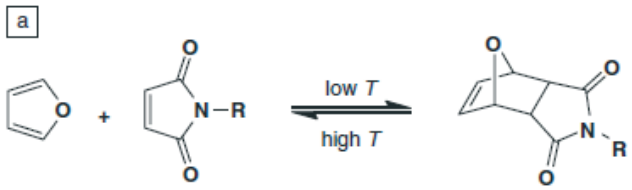


Fig. 3: *General reversible Diels-Alder scheme*

- **Di-Sulfide Links**:

The use of allylic disulfide linkages in self-healing materials has become popular due to their reversible cleavage [2,6]. The mechanism can involve radical or anionic intermediates and it is assumed that under mechanical load stress is locally released by cleavage of the bonds whereafter the system re-assembles and adopts a new stress-free state. The speed of reaction is strongly related to the unsaturation in the beta-carbon and is reported to be ineffective with aliphatic disulfides. A scheme showing the formation of the bonds is given in Fig. 4. It is clear that a combination of several mechanisms –in the example additional hydrogen-bonding (Fig. 5) - simultaneously yields the next generation of self-healing materials [5].

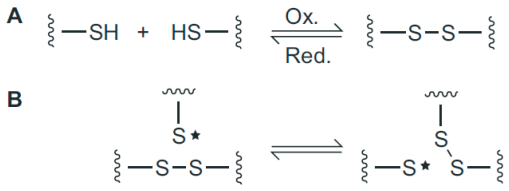


Fig. 4: *Thiol-disulfide reaction involving a radical intermediate*

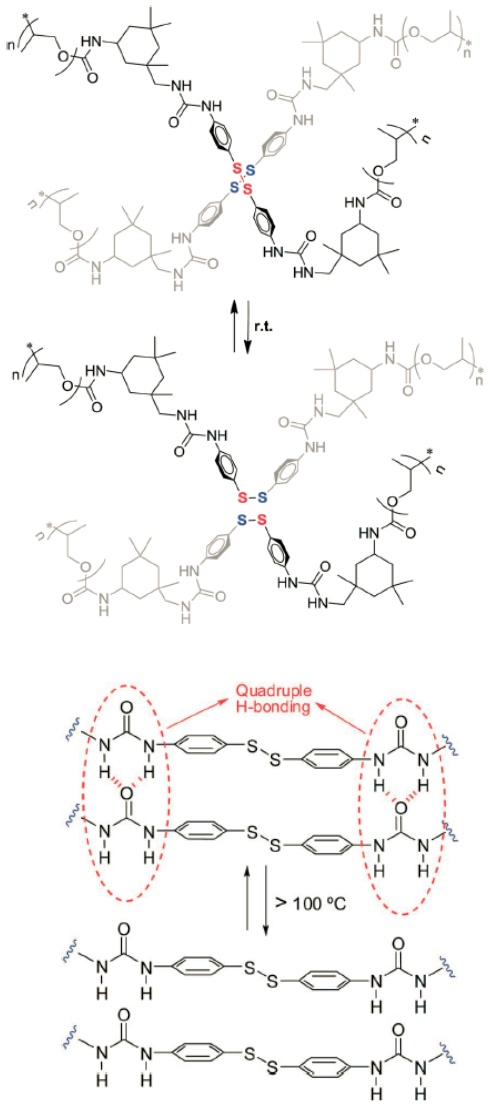


Fig. 5: *Example of the dynamic covalent disulphide linkage (top). Combination of a covalent linking with a supramolecular hydrogen bonding reaction, used in the 'terminator' polymer [5]*

2.3 Dynamic Non-Covalent Bonding-Based Self-Healing

In comparison to the above mentioned covalent counterparts, dynamic non-covalent bonds are in a continuous equilibrium and due to their weaker nature, they are more affected by factors such as thermal fluctuations, concentration, and solvent. Nevertheless, robust self-healing polymeric materials have been prepared using this type of interaction. The issues to deal with are primarily the time-scale of the dynamics, especially in relating the bond dynamics to the time scale of healing and also the strength of the formed associates.

- **Supramolecular Chemistry H-bonds:**

Hydrogen bonds play a prevalent role in functional biomaterials. Hence, hydrogen bonding has been extensively studied in supramolecular chemistry, both from a fundamental and an application driven viewpoint [3,4,6,8,9]. These prinicples are adapted in self-healing biomimetic polymeric materials .

In a dilute solution of telechelically modified polymers the polymerization degree (DP) strongly depends on the concentration, [M], of single supramolecular units and their respective equilibrium binding constants, K_a, following $DP \sim \sqrt{K_a[\mathrm{M}]}$ [2,3].

Long supramolecular polymer chains can only be achieved through multiple complementary hydrogen bonding interactions (Fig. 6). As the list of references and applications is very long we focus here on two approaches. Famous work was performed by the Meijer group in Eindhoven, using the UreidoPyrimidone (UPy) unit [3]. This has led to various applications in the field of supramolecular polymers. The general scheme is following:

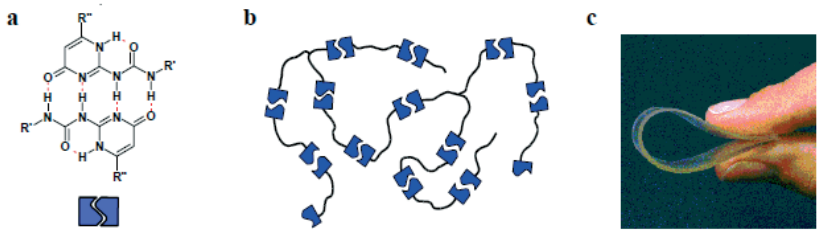


Fig. 6: *Quadruple Ureidopyrimidone-linker, linear aggregation and resulting supramolecular rubber [3]*

A further boost was given recently by the Leibler group in Paris [4]. Here, the components are aliphatic dimers and trimers of commercial fatty acids which, combined with a triamine and urea produced a material which is network-like and entirely builds upon a high number of hydrogen bridges.

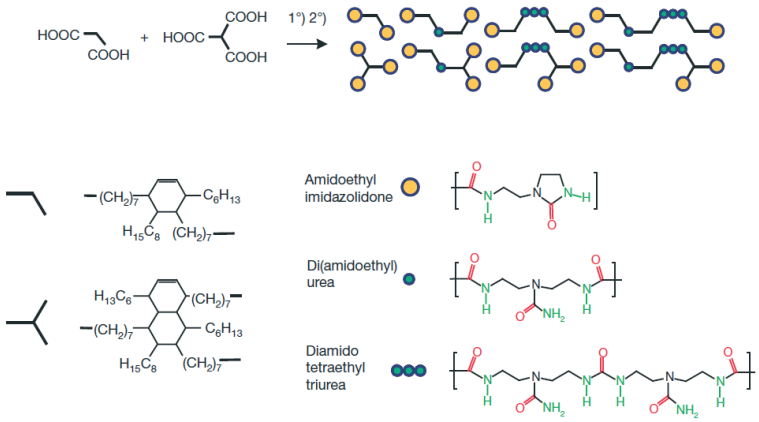


Fig. 7: *Natural fatty acids, combined with urea lead to a fully dynamic network [4]*

This rubber is amorphous at room temperature and dissolves its network upon temperature increase. The bonds open and the material can be reshaped (Fig. 7). Together with the high stability of the intermolecular bonds also a strong self-healing effect is found : fresh-cut surfaces intermolecularly diffuse and after 3 h the original Young modulus is restored! The efficiency of self-healing is matter of definition and treated later. A noted disadvantage is that the healability is retained only for a short time after the fracture event. Thermodynamic more stable structures may have formed and prohibit a new association.

- **Supramolecular Chemistry: π−π:**

Self-healing materials based on aromatic **π−π-stacking** interaction have been of interest as a convenient counterpart for the above mentioned hydrogen bonding mechanism [6,10,11]. A certain advantage is the insensitivity to humidity which destroys supramolecular H-bonding and saturates the supramolecular interactions with water molecules. Burattini reported the combination of **π-electron-rich and π-electron-poor moieties**. Fig. 8 illustrates how two π-electron-poor units which are connected by a flexible 'hinge' are able to back-fold and form a sandwich with a π-electron-rich donor by which a stacking situation arises. This aggregation modus typically shows high equilibrium constants. To overcome the rather high T_g of such materials heat is needed to start the self-healing processes. The strength and directionality of the interactions can be greatly enhanced by tailoring the electronic properties of the two different aromatic components.

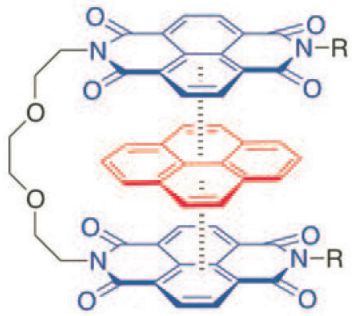


Fig. 8: *Example of a stacking geometry*

- **Metal-Ligand Coordination:**

For the design of supramolecular networks based on metal-ligand interactions, Pd and Pt in connection with suitable ligands are most often selected [6]. In order to create supramolecular networks the ligand requires at least two mono- or polydentate functionalities which are linked by a spacer group. Furthermore, the choice of the metal ranging from main- to transition-group metals to lanthanides and the oxidation state influences the stability of the generated complex and thus the dynamics as well as the reversibility of the metallo-supramolecular interaction. For this class, weak reversible interactions between the transition metals and e.g. aromatic amines like pyridine are chosen. The generic scheme is shown in Fig. 9 from which it is clear that the system can also adapt reversibly to stress. The method is not yet very popular and suffers from the presence of the metallic component. On the other hand, it is insensitive to moisture.

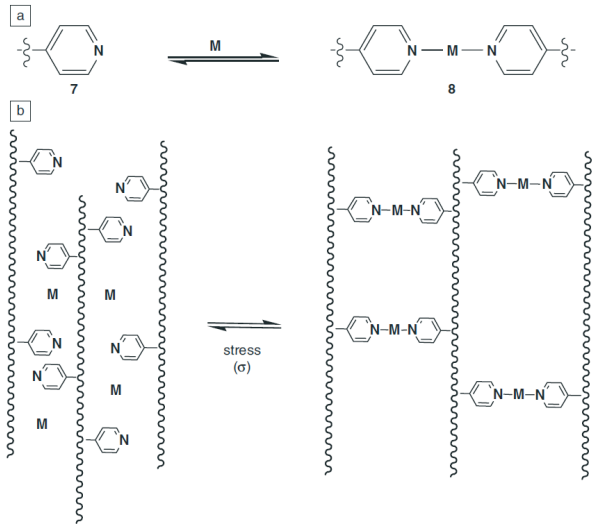


Fig. 9: *Exemplary pyridine-metal interaction and use in dynamic networking [6]*

- **Ionomers:**

The metal-ligand coordinative structures differ from so-called ionomers in that there is no charge involved [6,11]. Ionomers, on the other hand, are polymers which carry an appreciable

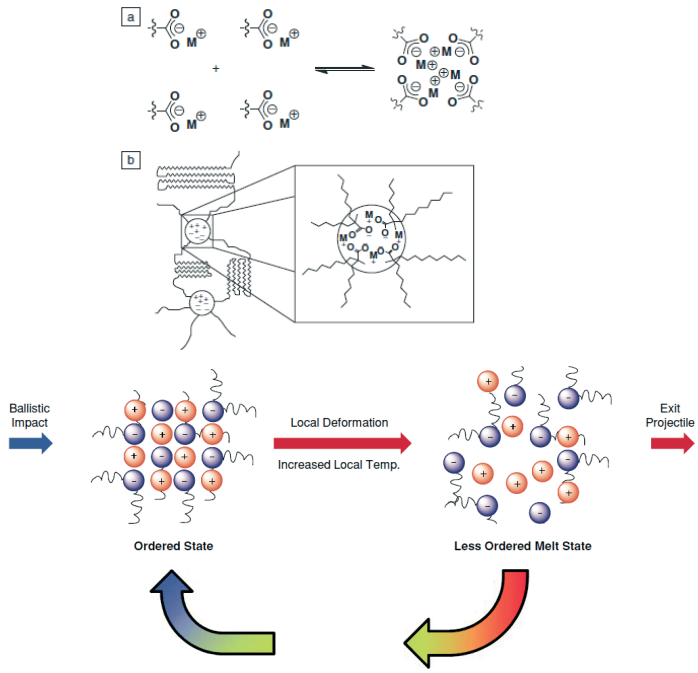


Fig. 10: *Ionomer composition and schematized ballistic self-healing*

amount of ions. Typical ionomers have grafted arms with acidic functions which are neutralized by a base. In this way separated ion-pairs result which form aggregates. The strong similarity with block copolymer micro-phase separation results from the fact that the charged groups tend to cluster into domains, and therefore charges are shielded from the amorphous matrix. The interactions are of electrostatic nature. The healing effect with ionomers has been demonstrated e.g. after a ballistic impact. In this case the energy required to facilitate self-healing is provided by the projectile. The process sets in due to the activation during high energy impact. This is an example where the type of damage and damage conditions must work in a synergistic manner with the polymeric structure.The structure of an ionomer and the self-healing process is summarized in Fig. 10.

3 Physical Characterization and Efficiency

Although not discussed in the former sections, one-time healing mechanisms have the strong disadvantage that a renewed microcrack at the same spot cannot be closed anymore. Intrinsic self-healing materials on the other hand can always be healed multiple times as the self-healing process and therefore the recovery of the materials properties are inherent to the system. A number of different methods have been developed to characterize and validate self-healing ablities that are based on different mechanisms and result in different material properties. These methods are rooted in engineering science and are based on complex theories of damage and fracture mechanics that will not be reviewed here [1,7,8]. Instead, this section will enumerate some definitions for the efficiency of the process and try to make a link from the macroscopic testing towards meso- or microscopic scales. Therewith, fundamental theories dealing with polymer dynamics and rubber elasticity can enter in the description and even constitutive models can be developed. Of course, all physical self-healing approaches like intermolecular diffusion or melting can also be considered as multiple time healing processes.

Healable materials are expected to regain their strength after a damage event. In Fig. 11 the tensile modulus of a system subject to consecutive break/heal cycles is shown. The efficiency of repeated repairing clearly depends on the material itself and the cycles may have several effects on the material. Thus, ultimatively, the performance of a sample will degrade, but, depending on the timescales involved, this may be adequate to fulfill the typical lifecycle of the product.

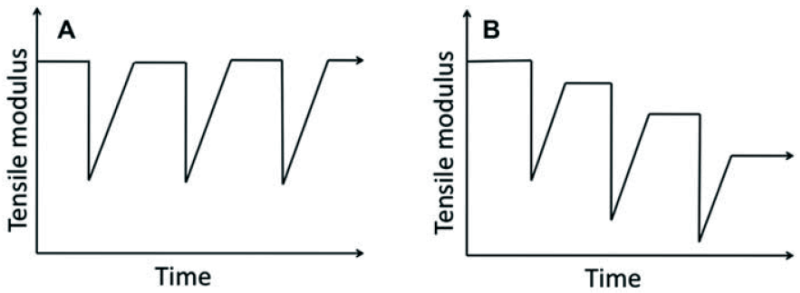


Fig. 11: *Ideal and real material strength after repeated self-healing cycles*

We need to quantify the loss in performance of the healed material. This is expressed in the '**healing efficiency**', $R(f)=f_{healded}/f_{initial}$, determined by the ratio of a specific mechanical property, f , of the material after healing and the initial state. Healing efficiency is therefore an easy parameter to obtain but the simplicity is misleading. A cross-correlation with different techniques is advantageous. In a typical break/heal experiment, a variety of mechanical properties for f are at hand: f=(tensile modulus, elongation–at-break, stress-at-break, modulus of toughness). At the ultimate tensile strength the material breaks. The energy required to break the sample is related to the area under the stress-strain curve and is referred to as the modulus of toughness.

Mechanical methods are most commonly used to detect and quantify the repair of a material. In a static experiment a constant force is applied to a pre-cracked sample which is typically shaped as a tapered double cantilever beam. The distance between the precrack and the spot where the force is applied is followed (Fig. 12). If the load is removed and the crack is allowed to heal through one of the above mentioned mechanisms, the crack length will again reduce and the ratio of the lengths defines the efficiency. For elastic samples, this geometry is not suitable and dog-bone shapes are chosen. The sample is not precracked. Instead, the sample is brought to rupture in the middle, whereafter the surfaces are allowed to interdiffuse, interreact and self-heal. From the comparison of the stress vs strain curves the relevant parameters can be deduced.

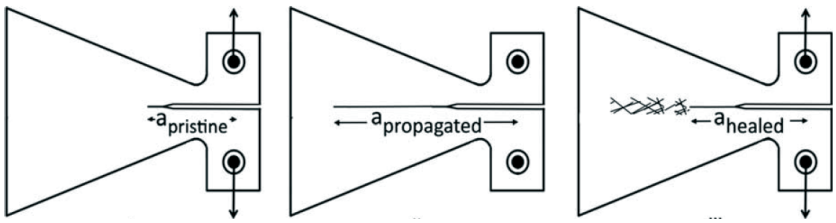


Fig. 12: *Double cantilever beam, crack propagation and subsequent healing [8].*

Instead of this monotonic static approach, also fatigue testing is common: here, the samples are tested oscillatory and the number of cycles of both virgin and healed samples before fatal rupture or reference value can be used for the efficiency measure.

From the scientific point of view the rheological technique is the most appropriate. The inherent transient behavior that self-healing polymers exhibit has already led to different models of rubber elasticity, where two different kind of bonds determine the elastic behavior. Polymers are viscoelastic materials that react elastically or viscous depending on the rate of perturbation or intermediate i.e. viscoelastic. The complex modulus then can be decomposed into a storage modulus G'(ω), corresponding to the elastic response to the strain and a loss modulus G"(ω) which is out-of-phase by 90° with the strain. Following both allows to obtain the increase of elasticity due to the re-association process and also to determine e.g. the time scales below which the dynamic bonds are more in a closed state than in the open state. Due to the similarity with networks, several approaches nowadays deal with the combination of both permanent and transient linkages. They base on the typical Maxwell and Kelvin-Voigt phenomenological models, i.e. spring and dashpot symbols to mimic the time-dependent properties of such systems [3,6,9]. Some assumptions are made : the permanent bonds are infinitely stronger than the transient ones, so constant over time, and the dissociation is faster

than the rebuilding process which has implications for the different rates. Without detailing further, but referring to the literature, already good accordance with creep data e.g. can be obtained. The constitutive model bases only on 2 different elastic moduli (one each for the permanent and the dynamic network) and a single viscosity which depends on temperature. The same concepts apply to shape-memory materials which form another new interesting class [9].

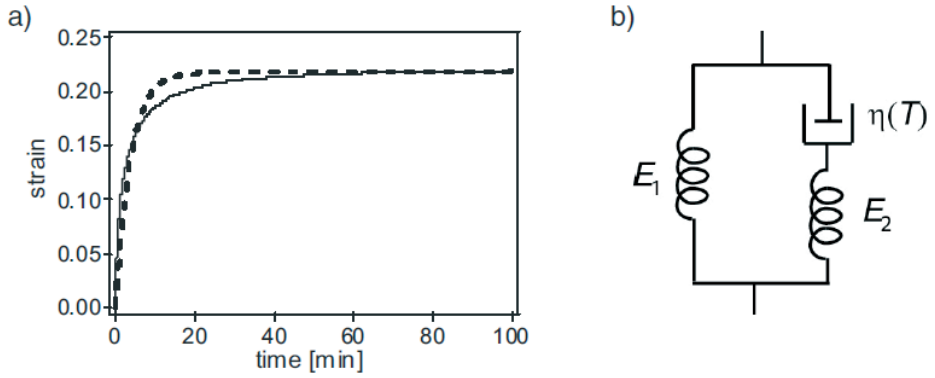


Fig. 13: *Creep measurement of a double network with covalent and non-covalent cross-links. Phenomenological spring-dashpot model describes the mechanical performance [9].*

The Leibler example in the previous section [4] clearly shows such a polymer-like behavior although the components are all based on small molecules (Fig. 14). At high frequencies the dynamical glass transition is seen, whereas at the lowest frequencies a frequency-independent elastic rubbery plateau in G' is shown. The existence of a transient rubber is further confirmed in uniaxial deformation. Stress–strain curves resemble those of soft rubbers and the strain at break exceeds 500% (Fig. 15). The residual strain after stress removal is less than 5%. When the cycle is repeated, the sample recovers completely without any residual strain. As for real rubbers the deformation follows incompressibility.

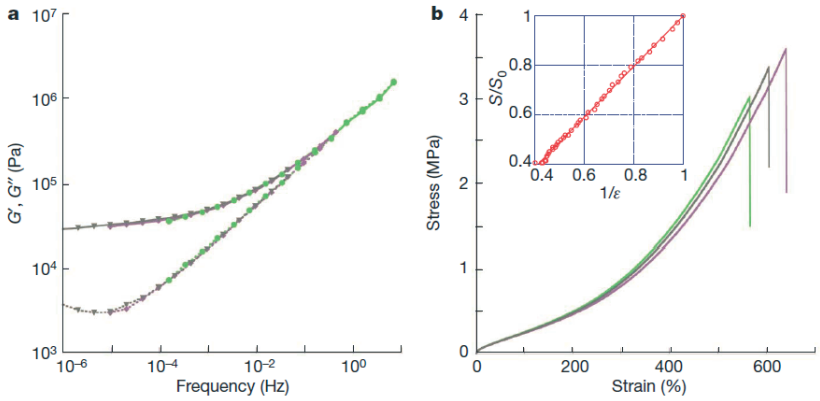


Fig. 14: *Formation of the supramolecular network in a frequency-independent storage modulus (left) and its similarity to real covalent networks (right) [4].*

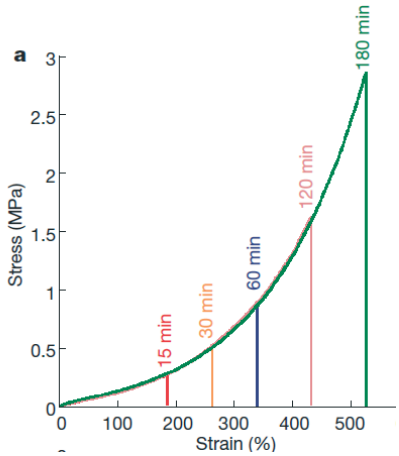


Fig. 15: *Self-healing test of the fatty acid mixture after joining fresh cut surfaces as a function of the healing time [4].*

4 Future Directions

It is still a grand challenge to meet all of the posed aims. Nevertheless, utilizing reversible cross-links of whatever kind within a polymeric network results in a new class of smart materials that can ultimately enlarge the service lifetime of polymeric materials for a wide range of advanced materials applications. Materials thereby receive an active role in target applications instead of being mainly passive. Prolonged lifetimes increase the reliability and enhance safety in cars and machinery and transportation infrastructure. Societal benefits are recognized by reducing the number of accidents, injuries and fatalities, the environmental pollution and urban noise. Economic benefits include less maintenance, less traffic jams and waste of time, savings in energy and natural resources consumption, reduced machinery idle-time due to frequent reparations, and reduced transportation costs.

To conclude, this review had the aim to guide the reader through the strongly expanding, interdisciplinary field of self-healing polymers which draws inspiration from physics, chemistry, biology and engineering. We tried to cover the fundamental principles that guide the development of such self-healing polymers, while many more ideas, techniques and applications are currently emerging.

References

[1]S. Van der Zwaag, Self Healing Materials - An Alternative Approach to 20 Centuries of Materials Science (Springer, Dordrecht, 2007)
[2]W. Binder and R. Zirbs, Adv. Polym Sci 207, 1 (2007)
[3]A.W. Bosman et al, Macromol. Symp, 201, 143(2003)
[4]P. Cordier, F. Tournilhac, C. Soulie-Ziakovic, L. Leibler, Nature, 451, 977(2008)
[5]A. Rekondo et al, J. Mater. Chem A, 2, 5710(2014)
[6]W. Binder, Self Healing Polymers: from Principles to Applications (Wiley-VCH, Weinheim, 2013)
[7]D. Wu, S. Meure, D. Salomon, Progr. Polym. Sci, 33, 479(2008)
[8]G. Wilson et al, Encyclopedia of Polymer Science and Technology (John Wiley and Sons, 2010)
[9]J. Li et al, Adv Mater, 19, 2851(2007)
[10]S. Burattini et al, Chem. Soc. Rev, 39, 1973(2010)
[11]S. Gosh, Self-healing Materials (Wiley-VCH, Weinheim, 2009)

G 2 Membranes for fuel cells and electrolysers

Uwe Reimer, Werner Lehnert
Electrochemical Process Engineering
Institute of Energy and Climate Research
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Polymer Membranes **4**

3 Electrolytes **7**

4 Structure of the membrane electrolyte composite **9**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Fuel cells and electrolysers are special applications of membrane reactors. The general setup can be explained by using the well known lemon battery as an example (figure 1). Here, two electrodes are supported and separated by the body of the lemon fruit. Lemon juice serves as electrolyte and provides ionic conductivity.

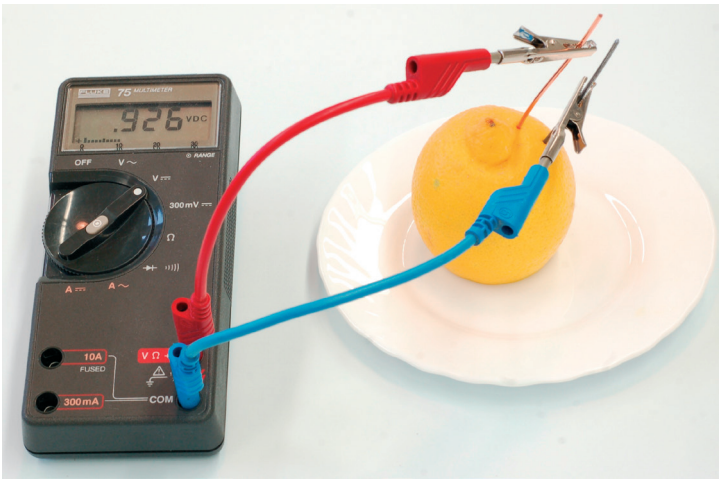


Fig. 1: *The lemon battery*

The energy delivered by the lemon battery comes from the dissolution of an ignoble metal while a noble metal serves as counter electrode.

$$\begin{aligned} &\text{anode:} && 2\,\mathrm{Al} && \rightarrow\ 2\,\mathrm{Al}^{3+} + 6\,e^- \\ &\text{cathode:} && 6\,\mathrm{H}^+ + 6\,e^-(\mathrm{Cu}) && \rightarrow\ 3\,\mathrm{H_2} \end{aligned}$$

In this case a 'biomembrane' supports and separates the two electrodes. Additionally, it also provides ionic conductivity since the juice is inside the membrane. The same is true for technical membranes where polymer membranes and electrolyte form a functional unit. The final behavior of membranes under working condition can only be understood or predicted if the characterization of the underlying materials is combined with an understanding of the three dimensional structure of the composite. The main focus of this chapter is to provide an overview over state of the art materials and their fundamental properties that lead to the successful application in fuel cells and electrolysers.

Figure 2 shows a sketch of a polymer electrolyte fuel cell. At the anode side hydrogen is supplied and at the cathode side either pure oxygen or air. Hydrogen molecules are oxidised to protons which are transported through the membrane to the cathode side whereas the electrons move through the extermal load to the cathode. At the cathode oxygen and protons form water molecules.

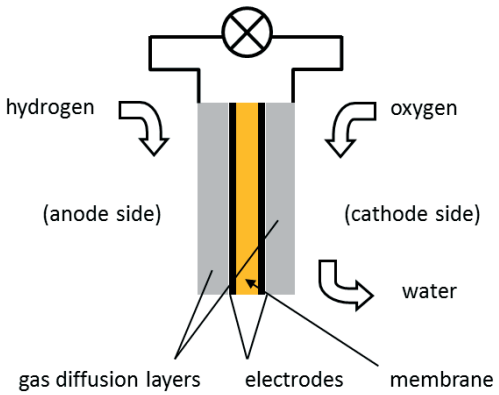


Fig. 2: *Sketch of a polymer electrolyte fuel cell*

The following scheme shows the respective half cell reactions at anode and cathode. Net reaction of a fuel cell is the combustion of hydrogen with oxygen to water. For electrolysers these reactions remain the same in reverse direction. The function of the membrane in these systems is to supply ionic conductivity, separate the electrodes and to block gas permeation from anode and cathode side.

$$\begin{aligned} \text{anode:} \quad & \mathrm{H_2} && \rightarrow \; 2\ \mathrm{H^+} + 2\,e^- \\ \text{cathode:} \quad & 0.5\,\mathrm{O_2} + 2\,\mathrm{H^+} + 2\,e^- && \rightarrow \; 0.5\,\mathrm{H_2O} \end{aligned}$$

The standard voltage for the reaction above is 1.23 V at room temperature, 101.325 kPa and water in its liquid state. The voltage of an operating fuel cell under load is usually between 0.5 V and 0.7 V because additional losses lead to a deviation from the thermodynamic value when electrical current is generated. In general three main effects contribute to losses. Firstly, kinetic losses at the electrode lead to a decrease of the cell voltage. This effect is usually called overpotential and refers to electron transfer processes. Secondly, the necessity of mass transport leads to additional losses. In some cases mass transport coupled with the electrochemical reaction solely generates concentration gradients which in turn change the overall voltage like described by the Nernst equation. Additionally, transport is always connected to some kind of resistance. The proton resistance (or conductivity) of the membrane is therefore one of the key parameters of a fuel cell or electrolyser which has a majour influence on the overall efficiency of the device.

2 Polymer Membranes

The main function of the polymer membrane is the spatial separation of electrodes while providing at the same time a matrix for the electrolyte. Therefore, the membrane should not have electronic conductivity but may possess ionic conductivity. It must have a high compatibility with the electrolyte, i. e. it must take up and contain electrolyte easily. For a technical application the membrane should also show high chemical and thermal stability. The operation temperature ranges from 25 °C to 200 °C in either strongly acidic or strongly alkaline environment. Furthermore, membranes should be stable under thermal cycling and insensitive to pressure gradients. Above all membrane materials are expected to be inexpensive and available on an industrial scale. The most successful candidates to withstand these drastic conditions are polymers that contain perfluorinated chains and/ or aromatic groups. At present there is a vast number of polymer and polymer blends available for application in electrochemical membrane reactors [1–5]. This text will give only few examples of prominent polymers.
The most established membrane is Nafion with the chemical structure shown in figure 3. It is one example of the large group of perflourosulfonic acid polymer membranes (PFSA). PFSA type membranes consist of carbon-flourine backbone chains with perflouro side chains that contain one or more sulfonic acid groups. Table 1 summarizes some of the commercially available Nafion types with the structure parameters a, b, c, d from figure 3 (as available).

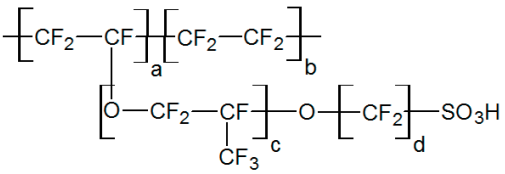


Fig. 3: *Repeating unit of PFSA type polymers (Nafion)*

Table 1: *Examples of commercial PFSA membranes [1]*

structure parameters				trade name	company	thickness / µm
a	b	c	d			
5 to 13.5	1	1	2	Nafion 120	DuPont	260
5 to 13.5	1	1	2	Nafion 117	DuPont	175
5 to 13.5	1	1	2	Nafion 115	DuPont	125
5 to 13.5	1	1	2	Nafion 112	DuPont	80
		0; 1	1 to 5	Flemion-T	Asahi Glass	120
		0; 1	1 to 5	Flemion-S	Asahi Glass	80
		0; 1	1 to 5	Flemion-R	Asahi Glass	50
1.5 to 14		0	2 to 5	Aciplex-S	Asahi Chemicals	25 to 100
3.6 to 10		0	2	Dow	Dow Chemical	125

One of the obvious differences among commercial membranes from table 1 are variations in lenght of the side chain and number of spacer groups in the backbone. An example of the three dimensional structure is shown in figure 4. The polymer chain contains a hydrophobic backbone and hydrophilic side chains. The final three dimensional structure is determined by the interaction of the polymer with water resulting in an ionic conductive system (polymer electrolyte mebrane). Other additivies may be present in order to increase hydrophilicity or mechanical strength of the system [1, 6]. Polymer and electrolyte form a bicontinuous structure as it is known from other complex fluids [7]. The driving force is similar to the concept of protein folding [8] although these membrane systems show less complexity. Nafion is likely to be the most frequently used polymer in polymer electrolyte fuel cells which operate at temperatures between 50 °C and 90 °C.

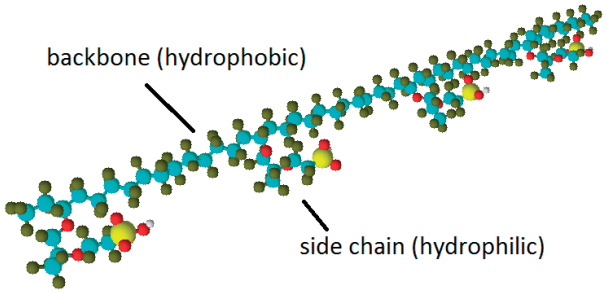


Fig. 4: *Ball and stick model of a polymer fragment for Nafion with structure parameters* $a = 5$*;* $b = 1$*;* $c = 1$ *and* $a = 2$

Aromatic hydrocarbons are another large class of polymers besides PFSA that show high chemical and thermal stability. Among them polyether ether ketones (PEEK) and sulfonated polyether ether ketones (SPEEK) are prominent candidates [9, 10]. The chemical structure of PEEK and SPEEK is shown in figure 5. The incorporation of sulfonic acid groups is frequently used to further increase hydrophilicity of polymer membranes. PEEK has been known under the brand name *Victrex* and shows very high stablity against oxidation at elevated temperatures but suffers from a comparably low glas transition temperature of 143 °C [1].

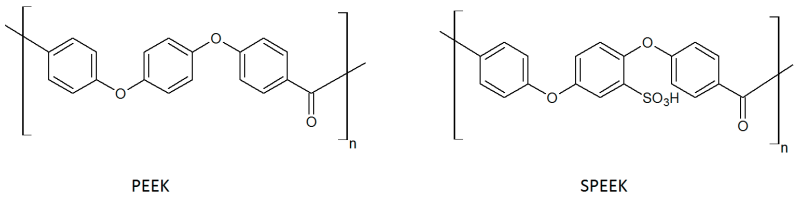


Fig. 5: *Repeating units of polyether ether ketones (PEEK) and sulfonated polyether ether ketones (SPEEK)*

Polybenzimidazoles are aromatic heterocyclic polymers containing benzimidazole units that also show superior stability against oxydation. These polymers have been developed originally for the production of textile fibres to be used in heat protection garments. They gained technical relevance for the application in fuel cells as they can be combined with phosphoric acid as an electrolyte. The most well-known polymer is poly 2,2'-m-(phenylene)-5,5'-bibenzimidazole (PBI or mPBI) with its brand name *Celazole* [3]. Another example with technical importance is poly(2,5-polybenzimidazole) (AB-PBI) [3]. The chemical structures of mPBI and AB-PBI are shown in figure 6. AB-PBI has a simpler structure and can be polymerized from a single monomer, which makes it potentially cheaper than mPBI. PBI type membranes are used in combination with phosphoric acid in high temperature polymer electrolyte fuel cells which operate at temperatures between 150 °C and 180 °C.

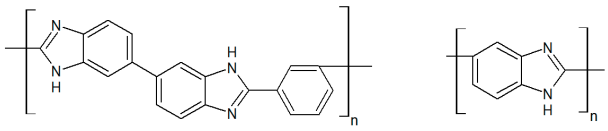


Fig. 6: *Repeating units of PBI (mPBI) and AB-PBI [3]*

Recently, polymer membranes based on graphene have attracted some attention. Graphene shows extraordinary electronic properties [11] but it can also serve as a virtually two dimensional polymer backbone. For this purpose graphene is first oxidized into graphene oxide. This can be used directly as self supporting membrane [12]. Another possibility is to introduce sulfonic acid groups in subsequent steps leading to an overall structure similar to the sketch in figure 7. In that case the final membrane is a composite of sulfonated polyimide and sulfonated propylsilane graphene oxide (SPI-SPSGO) and it was successfuly tested in a direct methanol fuel cell [13].

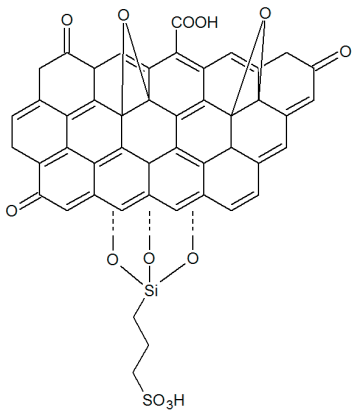


Fig. 7: *Sketch of the structure of sulfonated propylsilane graphene oxide (SPSGO) [13]*

3 Electrolytes

Aqueous solutions are the most common electrolytes for low temperature fuel cells and electrolysers. A special feature of these technological applications is that water also takes part in the electrochemical reaction. Therefore, water content changes as a function of operation conditions. Presently, structure and molecular properties of water are not fully understood [8] although water plays a key role in many technological and biological processes. One particular property of water is the ability to form hydrogen bonds. This leads to a tetrahedral coordination of the oxygen atom in bulk water as sketched in figure 8 [14, 15]. Electron density can be exchanged very fast between covalent bonds and hydrogen bridge bonds. The result is a network of water molecules with a certain degree of order [14, 15]. This cooperative effect plays a major role for the difference in macroscopic properties especially if only small volumes are occupied by water. The interaction with surfaces additionally changes the structure of the water network.

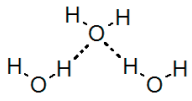


Fig. 8: *Sketch of hydrogen bonds in water*

The protonic conductivity of water is generally described by two different mechanisms. Firstly, protons can be transferred through the network of hydrogen bonds merely by the shift of local electron density. This leads to a very fast process where a proton is incorporated into the network at one location and another proton is simultanously released at another location. The largest time step in this process is the reorientation of water molecules. This process is sometimes called 'Grotthus mechanism'. Secondly, a so called 'hydronium ion' (H_3O^+) can be formed by autoprotolysis or by adding acidic substances. This ion can be transported by means of diffusion. Like all ions it will be surrounded by a hydrate shell. The characteristic time scale of this transport process will be similar to other diffusion processes in liquids. In acidic solutions exist a large variety of such 'hydronium ions' with the overall formula $H(H_2O)_n^+$ where only one of 10^{14} protons will have $n = 1$ [16]. The total protonic conductivity of aqueous solutions can be described by a mixed interpretation of the two mechanisms.

Under ambient conditions water can be used as electrolyte up to its boiling point at 100 °C. For some technological applications it is favorable to further increase the operation temperature. Reasons for this are a decreased sensitivity of the platinum catalyst to carbon monoxide poisening, easier exchange of heat for larger devices and the option to recover waste heat. Therefore, orthophosphoric acid (short: phosphoric acid) is used in some cases to increase the operation temperature up to 220 °C. Pure phosphoric acid has a melting point of 21 °C and theoretically a boiling point of 158 °C. Practically phosphoric acid has a very small vapor pressure. At higher temperatures water is released and phosphoric acid forms so called poly-acids [17] while still keeping a small amount of associated water. Generally, phosphoric acid has a very high affinity to water. It is therefore used as drying agent in the laboratory.

Phosphoric acid and water together also form a network of hydrogen bonds. In the case of pure water all protons are connected in the system of hydrogen bonds. For 85 % phosphoric acid at room temperature it was shown that ninety percent of all protons from acid and water are involved in such a network of hydrogen bonds [18]. (85 % phosphoric acid has a molar ratio of acid : water = 1 : 1.) Again, there are two mechanisms for protonic conductivity. Firstly, the so called 'Grotthus mechanism' seems to be a valid description just like in the case of pure water [18,19] (see figure 9). Secondly, phosphoric acid dissociates in three steps. Subsequently, the diffusion of 'hydronium ions' provides a second mechanism for proton conductivity which is now accompanied with the diffusion of the counter ions.

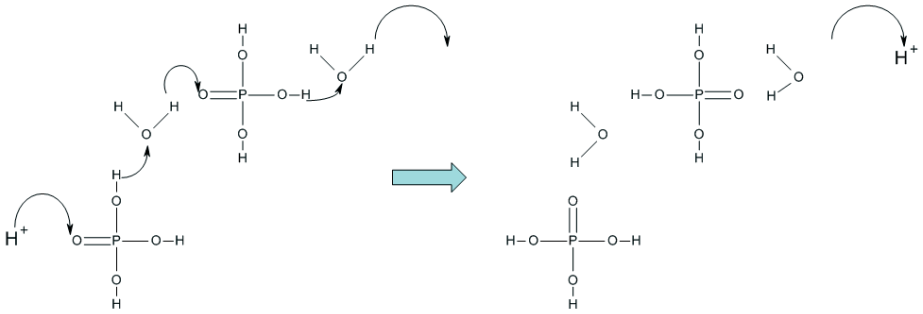


Fig. 9: *Sketch of the Grotthus mechanism of proton transfer in the system phosphoric acid/ water*

4 Structure of the membrane electrolyte composite

The morphology of the membrane is influenced by the structure of the polymer and its interaction with electrolyte and additives. The amount and composition of the electrolyte may change as function of the operating conditions. This leads to macroscopic swelling and shrinking of membranes, which lead in turn to differences in overall proton conductivity and also to mechanical stress and finally to degradation phenomena. Since polymer and electrolyte form a bicontinuous phase, the mechanical properties of the membrane is determined by the three dimensional structure of the (hydrophobic) domains of the polymer backbone. On the other hand, proton conductivity is provided by the domain of the electrolyte. The exact structure and dynamics of such membranes is very hard to access. First attempts have been made to simulate the overall structure for different water content with dissipative particle dynamics [20, 21]. The polymer forms a sponge like structure that contains a network of water filled pores. These simulations seem to be in agreement with data from small angle x-ray scattering [20]. The pore network of the electrolyte consists of clusters ('pockets') which are connected by channels. It is important to note that this electrolyte domain is not homogeneous in itself as shall be explained in the following (for water as the main electrolyte). The interface between water and polymer backbone consists of hydrophilic groups. These groups are hydrated, i. e. a layer of water is bound to these groups. The overall ordering of water molecules as well as their interaction differs significantly from bulk water. With increasing distance from the polymer surface the electrolyte shows more behavior like in the bulk phase. Nevertheless, the network always resembles a confined geometry where bulk properties do not necessarily apply. Figure 10 shows a two dimensional sketch of the assumed network of water pores.

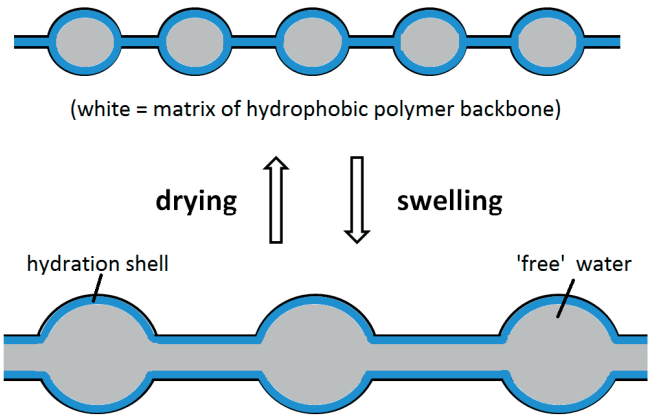


Fig. 10: *2D sketch of pores at low and high water content*

In figure 10 the water film of the hydrated polymer surface is shown in light blue and 'free' water in gray. With change of water content the membrane swells or shrinks. By doing so the size and number of water pores may change as well as the diameter of the connecting channels [1, 10, 22, 23]. This in turn should lead to differences in proton conductivity because the underlying mechanism depends upon the internal water structure.

The above described scenario was developed mainly through interpretation of macroscopic effects. Figure 11 shows the adsorption of water for PSFA polymers. With increasing partial pressure of water vapor the ratio of water molecules to surface groups n_{H_2O}/n_{SO_3H} changes very slowly, until at $n_{H_2O}/n_{SO_3H} \approx 5$ a sudden increase is observed. From this it can be deduced that on average five water molecules are required to hydrate one surface group. It is assumed that the structure of the PFSA membrane differs greatly from its dry state to its hydrated state [24]. Once the surface groups are completely hydrated, additional water molecules are incorporated as 'free' water. The final ratio for water adsorption from the gas phase seems to be $n_{H_2O}/n_{SO_3H} \approx 14$ and from the liquid phase $n_{H_2O}/n_{SO_3H} > 20$ [25].

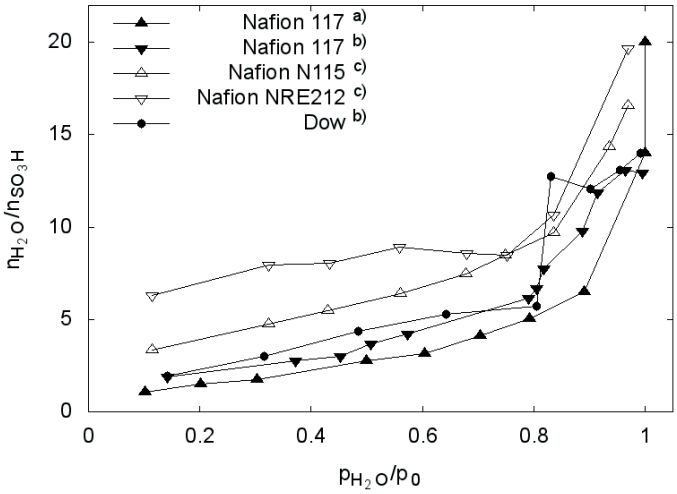


Fig. 11: *Hydration of PSFA polymers as function of normalized water vapour pressure; data from a) [26] at 27°C, b) [25] at 30°C, c) [27] at 30°C*

The resulting proton conductivity of these membranes is shown in figure 12. For values of $n_{H_2O}/n_{SO_3H} < 5$ the conductivity is almost zero. In this region all water molecules belong to the hydration layer. For larger values of n_{H_2O}/n_{SO_3H} conductivity increases. The slopes and maximum values in figure 12 depend on the inner structure of the polymer electrolyte composite as well as on mass transport issues in the respective electrochemical device.
The movement of protons from anode to cathode leads in turn to movement of water molecules in the same direction because of the attached hydrate shell. This phenomenon is called *electroosmotic drag*. For PSFA type membranes a number of 1 to 4 water molecules per proton are transported by this mechanism [25]. The resulting gradient in water concentration causes back diffusion of water from cathode to anode. Thus, the flux of protons is coupled to a flux of water molecules in operating fuel cells and electrolysers which in turn is strongly depending on cell geometry and operating conditions.

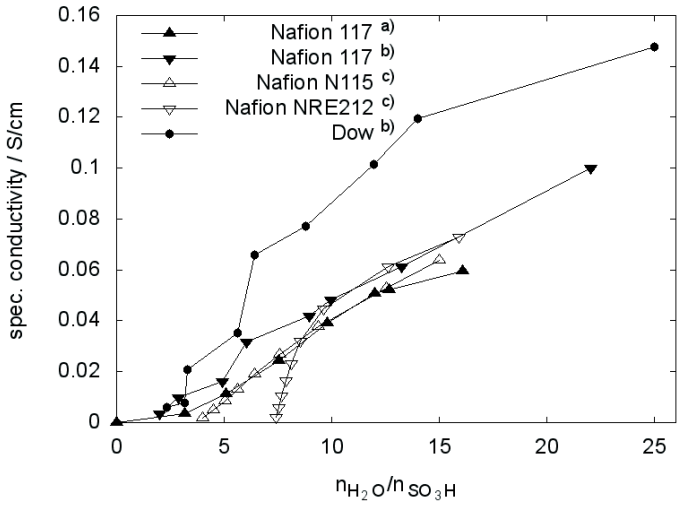


Fig. 12: *Proton conductivity of PSFA polymers as function of water content; data from a) [26] at 27 °C, b) [25] at 30 °C, c) [27] at 30 °C*

Today, the interpretation of small angle x-ray and neutron scattering data allows a more detailed analysis of the membrane electrolyte composite. For Nafion membranes a model was published that explains the characteristic spectra by an arrangement of long parallel water channels which show a high degree of short range order [28]. The membrane was characterized at a very low water content of 20 vol%. The resulting water channels have diameters in the range of 3 nm. The polymer backbone forms elongated crosslinks with an average cross section of 5 nm^2.
In order to predict the macroscopic conductivity of membranes it is necessary to understand how proton conductivity works in such small geometries with strong adsorption effects. To answer this questions it is necessary to perform neutron scattering experiments. Recent results for the system PBI/ phosphoric acid/ water show a distribution of proton diffusion constants with higher absolute values than anticipated from meassurements of macroscopic membrane conductivity [29].

References

[1] Q. F. Li, R. He, J. O. Jensen, and N. J. Bjerrum, Chemistry of Materials **15**, 4896 (2003).

[2] M. A. Hickner, H. Ghassemi, Y. S. Kim, B. R. Einsla, and J. E. McGrath, Chemical Reviews **104**, 4587 (2004).

[3] Q. F. Li, J. O. Jensen, R. F. Savinell, and N. J. Bjerrum, Progress in Polymer Science **34**, 449 (2009).

[4] S. J. Peighambardoust, S. Rowshanzamir, and M. Amjadi, International Journal of Hydrogen Energy **35**, 9349 (2010).

[5] S. Bose, T. Kuila, T. X. H. Nguyen, N. H. Kim, K. Lau, and J. H. Lee, Progress in Polymer Science **36**, 813 (2011).

[6] D. J. Kim, M. J. Jo, and S. Y. Nam, Journal of Industrial and Engineering Chemistry **in press** (2014).

[7] G. Gompper, in *Scattering Methods for Condensed Matter Research: Towards Novel Applications at Future Sources / Lecture Notes of the 43rd IFF Spring School 2012*, edited by M. Angst, T. Brückel, D. Richter, and R. Zorn (Forschungszentrum Jülich GmbH, Jülich/ Germany, 2012), p. B2.1.

[8] P. Ball, Nature **452**, 291 (2008).

[9] S. M. J. Zaidi, S. D. Mikhailenko, G. P. Robertson, M. D. Guiver, and S. Kaliaguine, Journal of Membrane Science **173**, 17 (2000).

[10] K. D. Kreuer, Journal of Membrane Science **185**, 29 (2001).

[11] M. I. Katsnelson, Materials Today **10**, 20 (2007).

[12] T. Bayer, S. R. Bishop, M. Nishihara, K. Sasaki, and S. M. Lyth, Journal of Power Sources **272**, 239 (2014).

[13] R. P. Pandey, A. K. Thakur, and V. K. Shahi, Applied Materials and Interfaces **6**, 16993 (2014).

[14] M. Dömer, T. Spura, R. Z. Khaliullin, and T. D. Kühne, Nachrichten aus der Chemie **61**, 1203 (2013).

[15] G. Malenkov, Journal of Physics: Condensed Matter pp. doi:10.1088/0953–8984/21/28/283101 (2009).

[16] V. Radtke, D. Himmel, K. Pütz, S. K. Goll, and I. Krossnig, Nachrichten aus der Chemie **62**, 1061 (2014).

[17] A. F. Holleman and N. Wiberg, *Lehrbuch der Anorganischen Chemie* (de Gruyter, Berlin, New York, 2007), 102nd ed.

[18] S. H. Chung, S. Bajue, and S. G. Greenbaum, Journal of Chemical Physics **112**, 8515 (2000).

[19] L. Vilciauskas, S. J. Paddison, and K.-D. Kreuer, Journal of Physical Chemistry A **113**, 9193 (2009).

[20] S. Yamamoto and S. Hyodo, Polymer Journal **35**, 519 (2003).

[21] H. Hayashi, S. Yamamoto, and S. Hyodo, International Journal of Modern Physics B **17**, 135 (2003).

[22] T. D. Gierke, G. E. Munn, and F. C. Wilson, Journal of Polymer Science B **19**, 1687 (1981).

[23] K. Jiao and X. Li, Progress in Energy and Combustion Science **37**, 221 (2011).

[24] G. Gebel, Polymer **41**, 5829 (2000).

[25] T. A. Zawodzinski Jr., T. E. Springer, F. Uribe, and S. Gottesfeld, Solid State Ionics **60**, 199 (1993).

[26] K. D. Kreuer, Solid State Ionics **97**, 1 (1997).

[27] L. Maldonado, J.-C. Perrin, J. Dillet, and O. Lottin, Journal of Membrane Science **389**, 43 (2012).

[28] K. Schmidt-Rohr and Q. Chen, Nature Materials **7**, 75 (2008).

[29] O. Holderer, O. Ivanova, B. Hopfenmüller, M. Zamponi, W. Maier, A. Majerus, W. Lehnert, M. Monkenbusch, and R. Zorn, International Journal of Hydrogen Energy **39**, 21657 (2014).

G 3 Polyelectrolyte membranes

O. Holderer
Jülich Centre for Neutron Science (JCNS @ MLZ)
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Fuel Cell Membranes **2**

3 Techniques **3**
3.1 Neutron Scattering 3
3.2 Imaging techniques 5
3.3 Other techniques 5
3.4 Computer simulations 6

4 Interpretation of scattering experiments **7**
4.1 Structure 7
4.2 Dynamics 9

5 Microscopic structure of a proton exchange membrane **9**

6 Microscopic proton transport inside the membrane **10**
6.1 Conductivity and local proton transport 12

7 Summary **13**

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved.

1 Introduction

Fuel cells convert chemical energy in the form of a fuel such as hydrogen gas or alcohol into electrical energy. Many different types of fuel cells exist, varying in output power, operating temperature, fuel, size and weight [1]. The common principle of fuel cells is the separation of protons and electrons from hydrogen in the fuel, the protons are then transferred from the anode to the cathode side through a proton conducting membrane, while the electrons go via the electrical consumer to the cathode side of the fuel cell.
Important parameters in this context are the proton conductivity in the heart of the fuel cell, the proton conducting membrane, and the long term stability of the fuel cell. Details on the fuel cell design and operation are given in Chapter G2.
An understanding of the microscopic structure and mobility of all components is essential for improving the fuel cell design. The structure is relevant for the mechanical and chemical stability of the membrane. The path of the proton is affected by the local structure on length scales of the order of nanometers to micrometers. This contribution deals with the structure of polymer electrolyte fuel cell membranes on atomistic length scales up to mesoscopic length scales (some 100 nm), and on measurements of the proton mobility on such length scales which is closely related to the macroscopic proton conductivity.

2 Fuel Cell Membranes

Basically two kinds of fuel cells using polymer electrolyte membranes as proton conducting membrane are presently available [2]: Fuel cells operating at temperatures below 100 °C, most commonly relying on Nafion based membranes, and high temperature polymer electrolyte fuel cells (HT-PEFCs), which operate at temperatures of 140-180 °C. More details on fuel cell operation are given in Chapter G2.
Here we are interested in the local structure of the core of such a fuel cell and its relation with stability, conductivity and proton diffusion.
The most commonly used material in low temperature polymer electrolyte membranes for fuel cells is Nafion, a sulfonated flouropolymer with ionic properties, therefore also called ionomer. It has basically a fluorocarbon backbone (Teflon) with sidegroups containing sulfonate groups (SO_3H). The water contents plays a crucial role in these materials and provides the conduction properties [3, 4]. Ionomer groups are the essential ingredients for providing the proton conductivity, the backbone polymeric material can be varied. It has been reported in Ref. [5] that styrene-ethylene copolymers are candidates for proton conducting membranes.
High temperature PEFCs have most often a polymer structure such as polybenzimidazole, which is swollen with phosphoric acid for providing the required proton conductivity. More details on membrane properties and materials are given in Chapter G2.
The essential function of the polymer electrolyte membrane in fuel cell applications is the transport of protons across the membrane from the anode to the cathode. The environment where this proton transport takes place is rather complex. Polymer domains, maybe partly crystalline, partly amorphous [3], water, acid or ionomer containing parts are part of the membrane. There have been two main mechanisms proposed in the literature how the proton moves across the membrane, depicted in Figure 1. The carrier mechanism describes the situation where protons use e.g. H_2O as a vehicle and travel as a H_3O^+. The Grotthus mechanism describes the transport along hydrogen bonds by a combination of jumps and reorientation [6]. Transport in electrolyte

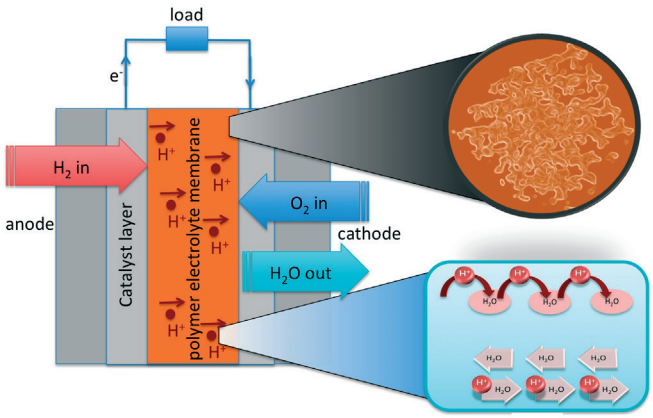


Fig. 1: *Left: Schematic representation of a fuel cell stack. H_2 gas on the anode side and O_2 from air at the cathode side serve at a fuel, the proton conducting membrane in the middle lets pass only H^+, the electrons take the way through the consumer as the current produced by the fuel cell. How to get information on the structure (top right) and the transport mechanisms (bottom right, e.g. hopping or carrier transport) will be discussed in this chapter.*

material plays a role not only in fuel cell applications, but also for batteries. Polyethylene oxide (PEO) is a solid polymer electrolyte used in Li-ion batteries as an ion conductor for Li^+[7]. The polymer dynamics and ion diffusion have been investigated with neutron scattering experiments [8]. Polymer dynamics also has been investigated in the context of Li-conduction in batteries with quasielastic neutron scattering in ref. [9].

3 Techniques

The goal is to achieve a good understanding of the structure and dynamics on microscopic length and time scales. Amongst the zoo of possible techniques the focus will be set here to only some selected techniques.

3.1 Neutron Scattering

Neutron scattering is excellently suited for studying soft matter samples due to the capability of contrast variation by isotop substitution. Introductory material on this topic can be found e.g. at neutronsources.org [10]. Structural properties can be adressed with small angle neutron scattering (SANS) and neutron diffraction, polymer dynamics and proton diffusion is the domain of quasielastic neutronscattering techniques, such as backscattering spectroscopy, time of flight spectroscopy or neutron spin echo spectroscopy [11, 12, 13]. There are different strong arguments for using neutron scattering. Absolute scattering intensities can be measured since the

scattering cross sections of the elements are known. SANS experiments allow therefore to obtain quantitative information on the sample, such as volume fractions of different components, the surface to volume ratio of two component systems, and by the dependence of the scattering intensity on the angle one obtains structural and shape information of the constituents of the sample. The irregular dependence of the scattering cross sections on the atomic number and on different isotopes allows to study also light elements. The high incoherent cross section of the proton makes proton diffusion studies possible. Neutrons pass through thick samples of the order of mm, it is therefore possible to study liquid samples thick samples. The drawback is that the results of scattering experiments are obtained in reciprocal space (i.e. a Fourier transformed image of the real space) and need some modeling of the data. Real space distances d and the modulus of reciprocal space scattering vectors q are related by $d = 2\pi/q$. The result is an averaged information of the sample, such as atomic position or particle size and not a view of single individual particles. The averaging over a macroscopic amount of sample requires that the sample needs to have a certain homogeneity w.r.t. the quantity one would like to observe. On the other hand the result is representative of the whole sample and not a pick at a single location as it is with microscopy techniques in real space.

Diffraction

SANS is a diffraction technique which gives structural information such as size and shape of nanoscopic objects. A highly collimated beam hits the sample, a detector behind the sample records the scattering at low angles (typically $< 10°$). This technique is (as its X-ray equivalent SAXS) sensitive to length scales of the order of nanometer to 100s of nanometers.

Neutron Diffraction is used to study atomic distances. In solids, the crystal structure can be obtained from single crystal diffraction or powder diffraction, but also in soft matter samples typical atomic repeat distances occur, e.g. the typical bond length of carbon atoms, which makes diffraction a useful tool also in disordered materials.

Spectroscopy

The momentum transfer during the scattering process tells us something about the structure of the sample. The energy transfer depends on the dynamics, i.e. the motion of the scattering particles in the sample. Depending on the required length and time scale, different spectroscopic techniques are available. Other than for three-axis spectrometers, where typically distinct excitations are measured, e.g. phonon dispersions in solids, the techniques described here are used in this context of "soft matter" to measure a continuous spectrum of energy transfers of thermally activated motions. This results in a broadening of the elastic line (the neutrons which did not suffer any energy transfer) and not to distinct peaks well separated from the elastic line. It is therefore often referred to as quasielastic neutron scattering (QENS or QNS).

Time of flight spectroscopy allows to study motions in solids on lengths scales below 1 nm and energy transfers in the meV range, corresponding to time scales of 1-100 ps. The neutron beam is chopped with a series of choppers such that a short pulse of known wavelength (=known speed) arrives at the sample. The travel time of the neutrons to the detectors in some meters distance is measured and deviations from the initial velocity are detected.

Backscattering spectroscopy is the technique with the highest resolution when measuring directly the energy transfer. The backscattering spectrometer consists, similar to a three axis spectrometer, of a monochromatizing crystal, the sample and analyzing crystals. The peculiarity here is that the Bragg-reflections at the monochromatizer and analyser take place in backscattering geometry (wavevector of the incoming neutrons is perpendicular to the surface). This minimizes the angular error in the Bragg equation and leads therefor to the best precision with crystal monochromatization. Energy differences of the incoming neutrons on the sample

are produced by moving the monochromator crystal and imposing a Doppler shift of the neutron energy. Energy transfers in the range of < 1 up to $\simeq$30 μeV corresponding to $\simeq$10-5000 ps are measured with this technique.
The large incoherent cross section of protons allows to study self diffusion processes of protons with neutron scattering.
A special case of backscattering measurements are so called elastic scans [13], where the evolution of the elastically scattered intensity with temperature gives information on the proton mean squared displacement in the sample on a time scale of some nanoseconds.
Neutron spin echo spectroscopy (NSE) measures the intermediate scattering function, the Fourier transform into the time domain of the scattering function S(q,ω) measured by TOF- or backscattering spectroscopy. Instead of measuring the velocity before and after the scattering event, only velocity changes are measured by a neutron spin encoding and decoding procedure. This gives the highest energy resolution in quasielastic neutron scattering, with time scales of up to 200 ns (energy range down to some neV). NSE spectroscopy can be used to study incoherent processes, but to the price of a large background since in incoherent scattering 2/3 of the beam gets depolarized and only 1/3 remains polarized for the velocity change analysis. On the other hand, NSE can be used to investigate the dynamics of coherently scattering parts of the sample, mainly in the region where SANS is used for structural investigations. In the case of polymer membranes, thermally driven fluctuations of the polymer chains can be investigated. Coherent contrast is achieved by deuterating the major part of the sample, the remaining protonated part can then be measured. For fuel cell membranes this means that deuterated solvent has to be used with the mainly protonated polymers (water in the case of Nafion or deuterated phosphoric acid for PBI membranes).

3.2 Imaging techniques

Electron microscopy provides real space images on the relevant nanoscopic length scales. The highes resolution is obtained with transmission electron microscopy (TEM), where the length scales match well those of neutron scattering experiments in reciprocal space. The requirements to the sample are completely different. A very thin slice of the sample of $\simeq 50$ nm is required to allow the electron beam pass through. Measurements take place in vacuum, i.e. the sample must be dry, or liquid samples need to be frozen and kept in this condition in a special sample holder (Cryo-TEM).
Scanning electron microscopy (SEM) looks at surfaces on length scales down to approximately 10 nm. The resolution is not as good as in the TEM, but sample preparation is much easier. Both techniques, TEM and SEM, can be equipped with chemical analysis options, such as energy dispersive x-ray spectroscopy (EDX) or electron energy loss spectroscopy (EELS), which allows to map chemical elements in the sample.
Optical microscopy allows for an inspection of the membrane on length scales of microns or above. Certain aging effects are probably reflected on length scales accessible by optical microscopy.
Atomic force microscopy is a surface sensitive technique, where one can achieve also very high resolution with appropriate tips.

3.3 Other techniques

Nuclear Magnetic Resonance

Pulse field gradient nuclear magnetic resonance (PFG-NMR) allows studying diffusion processes on intermediate length- and time-scales by following the spin polarization of e.g. hydrogen atoms. A radio frequency sequence of pulses creates a spin echo after an observation time τ. Gradient magnetic field pulses are applied at the sample, which define the length scale of observation (typically down to micrometers). The spin echo amplitude decays when particles diffuse along the magnetic field gradient due to the spacial dependence of the Larmor frequency in the gradient field. The spin echo attenuation measured in PFG-NMR experiments is a signal decaying from its initial value to zero, in analogy to the intermediate scattering function measured in NSE experiments. Only the time and length scales are rather in the micrometer/microsecond instead of the nanometer/nanosecond regime as in the NSE case. Both techniques are therefore complementary and allow to study diffusion processes over a large tempo-spacial range.
The local diffusion measurements can be linked directly to macroscopic properties such as the conductivity of the membrane and is therefore an important part of the puzzle in understanding the mechanisms of stability, performance and degradation in fuel cells.
Conductivity measurements as a macroscopic technique also provide valuable insight into the proton transport mechanisms, by linking the conductivity to that of pure materials, e.g. pure phosphoric acid. The comparison between macroscopic conductivity and microscopic diffusion can yield information on the nature of transport processes, for example which fraction of the proton motion actually contributes to the macroscopic conductivity.
Also many other techniques can be useful for the microscopic and macroscopic study of fuel cell membranes, such as dielectric spectroscopy, ^{1}H-NMR spectroscopy [14], X-ray and neutron radiography [15, 16], to mention just a few.

3.4 Computer simulations

Molecular dynamics (MD) simulations are a virtual experiment, where the trajectories of atoms or particles in general can be followed [17, 18, 19]. A possible free software for such MD simulations is gromacs[20]. The transport mechanism (Grotthus- and vehicle mechanism) has been simulated with ab initio MD simulations in Ref. [21]. The principle of ab initio molecular dynamics simulations is to write down the equation of motion for all atoms in the sample volume which is simulated:

$$m_i \ddot{\mathbf{r}}_\mathbf{i} = \mathbf{F}_\mathbf{i} \tag{1}$$

with the mass m_i and position $\mathbf{r_i}$ of each atom and the forces $\mathbf{F_i}$, which in the simplest case are taken as pairwise Coulombic interactions with an additional short range repulsion [22]. On each atom, the interaction forces of all surrounding atoms are acting, which is described with a suitable force field. In each step (=time step) of the simulation, the displacement of each atom according on the acting forces is calculated and the atom moved to the new position, then the procedure is repeated. A first "production run" of several time steps equilibrates the system and brings it into thermodynamic equilibrium (it produces a canonical ensemble NVT). Then the transport and dynamics of the atoms can be simulated. The limiting part of MD simulations is the box size, since each atom has to be treated individually, and the computation time for reaching simulation times corresponding to the typical time scale of neutron scattering experiments.
This can be overcome with coarse graining, i.e. several atoms are grouped together to a particle which is then observed in the simulations [23].
Since the atomic positions of particles can be tracked in MD simulations, it is easily possible to

obtain the structural or time correlation function from MD simulations which allows to link the simulations to scattering experiments.

4 Interpretation of scattering experiments

In chapter C1 correlation functions have been introduced and their relevance for static and dynamic scattering experiments has been pointed out. They describe the probability finding a system in state $A(t)$ when it was known to be in state $A(0)$ at $t = 0$ for some variable t (e.g. time or space). The spacial correlation function is called radial distribution function. Some aspects of the correlation function important for the characterization of polymer electrolyte membranes are pointed out in the following sections.

4.1 Structure

Scattering experiments measure static correlation functions. A very short introduction presents here how structural information are gathered from such experiments [24].
The density of scatterers in a sample can be written as

$$\rho(\vec{r}) = \sum \delta(\vec{r} - \vec{r_j}) \tag{2}$$

for point scatterers at the positions $\vec{r_j}$, where the sum runs over all scattering centers. The spacial or pair correlation function is then

$$\rho(\vec{r_1})\rho(\vec{r_2}) = \sum \delta(\vec{r_1} - \vec{r_j})\delta(\vec{r_2} - \vec{r_k}) \tag{3}$$

summing over all pairs of centres j and k. The Fourier transform of this equation is the structure factor measured by small angle scattering of neutrons or X-rays (with different contrast, i.e. different functions of the scattering density $\rho(\vec{r})$):

$$S(q) \propto \int d^3r \rho(\vec{r_1})\rho(\vec{r_2})e^{iq\vec{r}} \tag{4}$$

with $\vec{r_2} = \vec{r_1} - \vec{r_2}$.
Fractal structures exhibit a self similarity if looked at it on different length scales. This can be found in nature, e.g. in the structure of a cauliflower (see Fig. 2 [25]), but also on microscopic length scales, where the structure resembles itself over a rather broad range of length scales. Mass fractals have the property that the mass M scales as $M \propto r^D$ with the fractal dimension D. The density of a mass fractal structure has a length scale dependence of[26, 12]

$$\rho(\vec{r}) \propto \rho(\vec{r_1})^{D-d} \tag{5}$$

The Fourier transform is the scattering intensity in reciprocal space as measured e.g. by SANS or SAXS. For a fractal density distribution, the Fourier transform in polar coordinates leads to

$$I(q) \propto q^{-D} \tag{6}$$

i.e. when plotting the intensity vs. q on a double logarithmic scale, the slope is the fractal dimension. Since the mass of a homogeneous object scales as $M \propto r^3$, the upper limit for the mass fractal dimension is $D < 3$.



Fig. 2: *Cauliflowers: fractal structure made by nature.*

Porod law: Porod derived an expression for large scattering vectors q, which relates the scattering intensity to the ratio of surface S to volume V (see chapter C1 and Ref. [27]).

$$I(q) = \frac{2\pi}{q^4}(\Delta\rho)^2\frac{S}{V} \tag{7}$$

At large q, the intensity drops with q^{-4}, which is characteristic for sharp interfaces with a sudden variation in the scattering length density ρ. A rough or fractal surface results in a less strong q-dependence, with $I(q) \propto q^{-(6-D_s)}$ with the surface fractal dimension in the range of $2 < D_s < 3$[12].

Teubner Strey Model:

If density variations occur with a characteristic repeat distance d, a peak is visible in reciprocal space at a scattering vector $q_0 = 2\pi/d$. The more regular the structure is, the sharper gets the peak.

One possible way characterizing such a structure which exhibits a characteristic length scale but is nonetheless rather disordered is the Teubner-Strey model, where a Gaussian random field model is used to present the density fluctuations of the microemulsion. The scattering intensity

$$I(q) \propto \frac{1}{q^4 - 2(q_0^2 - \xi^{-2})q^2 + (q_0^2 + \xi^{-2})^4} \tag{8}$$

Eq. 8 shows a peak at $q_0 = 2\pi/d$ with a characteristic width proportional to the inverse of the correlation length ξ [28] .

Rodlike structures:

Rodlike objects have a typical small angle scattering intensity dependence of

$$I(q) = \frac{L\pi}{q}I_c(q) \tag{9}$$

which depends on the length of the rods, L, and the function $I_c(q)$ related to the rods cross section [27]. The q^{-1}-dependence of the scattering intensity has been interpreted as a rod like water channel structure in Nafion membranes [29, 3].

4.2 Dynamics

The van Hove correlation function $G(r,t)$ describes the probability finding a particle at time t at position r, if at time $t = 0$ another particle was at position $r = r_0$. If we look only at the self diffusion, e.g. of protons (N particles) in the case of fuel cells, the self correlation function is

$$G(r,t)_s = \frac{1}{N}\sum \delta(r - r_i(t) + r_i(t=0)) = \frac{1}{N}\sum \delta(r - \Delta r_i(t)) \tag{10}$$

For the time $t = 0$ we have $G(r, t = 0) = \delta(r - r_0)$, i.e. the system is in its initial state.
The Fourier transform of the correlation function into reciprocal space is the intermediate scattering function $S(q,t)$, in the case of self diffusion it reads:

$$S(q,t)_s = \frac{1}{N}\sum \langle \exp\left(iQ\Delta r_i(t)\right)\rangle \tag{11}$$

If the displacement of particles obeys a Gaussian distribution and we do a cumulant expansion, we can write $S(q,t)_s$ in the following form:

$$S(q,t)_s = \frac{1}{N}\exp\left(-\frac{1}{6}Q^2\langle \Delta r^2(t)\rangle\right) \tag{12}$$

The mean squared displacement (MSD) $\Delta r^2(t)$ directly measured with NSE spectroscopy or by Fourier transform from ω-space into the time domain of the scattering function $S(q,\omega)$ measured with TOF- or backscattering-spectroscopy (all instruments with their own characteristic time scale).
Diffusion has been introduced in chapter C1. As a short summing up the main result is repeated here. Starting from Fick's second law, the MSD has been linked to the diffusion constant D by

$$\langle \Delta r^2(t)\rangle = 2Dt \tag{13}$$

Microscopically, the movement of a particle in a random walk follows this equation. Figure 3 illustrates the evolution of the van Hove correlation function with time, i.e. a broadening of the particle distribution starting from a point where all positions have been known (sharp peak), which corresponds to a random walk of each particle.
The same result can also be obtained when starting from a microscopic point of view with the Langevin equation, the equation of motion of a particle in a fluid undergoing collisions with the fluid particles. The strong incoherent scattering cross section of the proton allows to measure the Fourier transform of the van Hove correlation function with QENS experiments (NSE, Backscattering and TOF spectroscopy). The MSD of the proton is therefor accessible in these experiments.

5 Microscopic structure of a proton exchange membrane

We have seen in the previous sections that neutron scattering techniques are a good choice for structural investigations of polymer electrolyte membranes. We will have a look now how

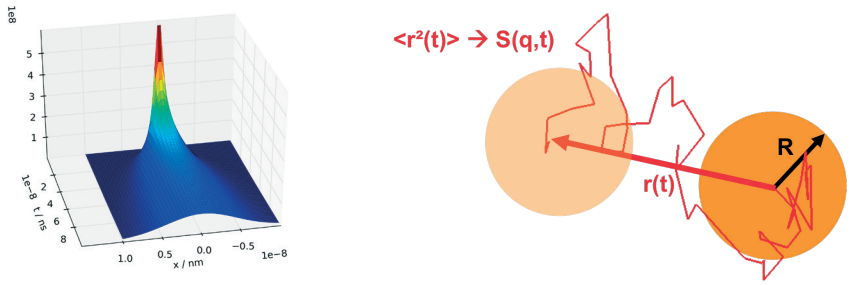


Fig. 3: *Left: At time t=0 the state of the system (all atomic positions) is known, all atoms are at their initial position. With evolving time, the atoms diffuse further and further away, they are distributed over a region with width σ from their initial positions. Right: If a particle moves away from the origin in a random walk, its distance covered at time t, $r(t)$, obeys the* $\langle \Delta r^2(t) \rangle = \langle r(t)^2 \rangle = 2Dt$.

the small angle scattering signal from a proton exchange membrane, in this case a sulfonated styrene-ethylene copolymer membrane [5] looks like. Figure 4 shows the SANS intensity as a function of scattering vector q, which is inversely proportional to the length scale investigated. The dry membrane shows some power law decay below $q = 0.01\mathring{A}^{-1}$ indicating a fractal structure over this length scales range (larger than $\simeq$ 50 nm). Hydrating the membrane, as it is needed for providing sufficient conductivity in fuel cell applications, the scattering significantly changes and a pronounced peak at $q = 0.1\mathring{A}^{-1}$ emerges. This peak has been interpreted with the Teubner-Strey model as a signature of a bicontinuous structure, with the polymer material as one phase and water channels as a second phase, both interpenetrating each other like a soaked sponge. From the fits to the Teubner-Strey model, characteristic length scales of the water domains and a correlation length, which indicates how ordered and regular the water domains are distributed, has been deduced. Figure 4 shows a possible picture of such a bicontinuous structure.

Similarly, Nafion membranes have been investigated with small angle X-ray scattering, which corroborated the picture of water channels due to a q^{-1}-dependent decay of the scattering intensity typical for rodlike structures, together with the appearance of a peak similar to the one in Figure 4 often called "ionomer peak", which always indicates a regular structure in the sample. The Nafion water channels are rather regularly structured. Figure 5 shows the small angle scattering intensity as it would be measured from a Nafion like membrane. References [3, 30] report on such measurements.

6 Microscopic proton transport inside the membrane

The proton diffusion as well as the polymer dynamics can be studied experimentally on a nanometric length scale with quasielastic neutron scattering techniques. Proton transport in pure phosphoric acid was investigated e.g. in Ref. [31, 32]. It has been shown that Nafion, one of the extensively studied materials, gets softer with increasing hydration level (plasticitation) up to a level where it is saturated with water [33]. The proton diffusion in hydrated Nafion membranes

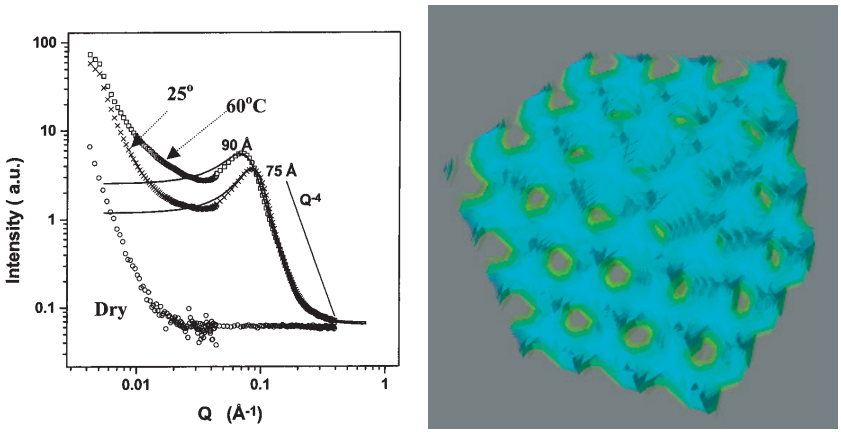


Fig. 4: *Left: Small Angle Neutron Scattering shows averaged structural properties on length scales of nm-μm. Reprint with permission from [5]. Copyright (2002) American Chemical Society. Right: Sketch resulting from the SANS/SAXS experiments, as described e.g. in [3]*

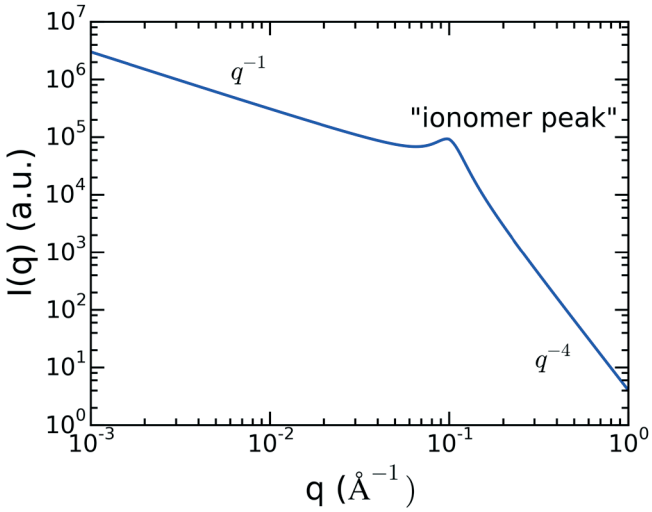


Fig. 5: *Small angle scattering (SANS or SAXS) from a Nafion membrane follows a q^{-1} power law at low q as a signature of cylindrical water channels as described e.g. in [3], the so called "ionomer peak" similar to the correlation peak from a structure as depicted in Fig. 4, and a q^{-4} Porod behavior at large q.*

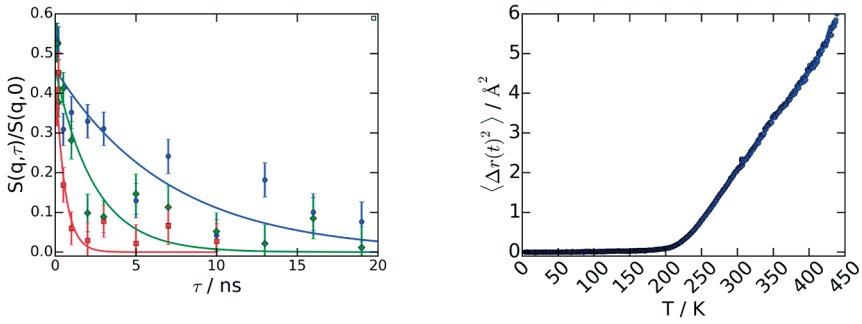


Fig. 6: *Left: Intermediate scattering function obtained from incoherent scattering of protons in a phosphoric acid loaded PBI membrane with NSE spectroscopy for 3 different q-values, with a simultaneous fit with a q^2-dependence of the relaxation rate. Right: MSD deduced from an elastic scan from a backscattering experiment on a PBI membrane loaded with phosphoric acid (see [35]).*

has been studied with QENS in Ref. [34]. They observed two populations of diffusing protons, a slow and a fast one, which they attributed to different diffusion processes. In Ref. [35] it has been shown that the polymer scaffold of a PBI membrane loaded with phosphoric acid as a proton conductor is very rigid, providing the needed stability of the sample. The intermediate scattering function in a NSE experiment, $S(q,t)$, is a constant if no fluctuations are present. On the other hand, incoherent scattering from the protons by NSE and backscattering revealed that the proton diffusion is only slightly lower than that of pure phosphoric acid. Elastic scans at a backscattering spectrometer and the NSE results already indicated that the diffusion is not governed by one single diffusion constant, but by a distribution of constants[35], since on different length scales the value of the diffusion constant varied significantly. Figure 6 show the intermediate scattering function from incoherent scattering measured by NSE on phosphoric acid loaded PBI membranes, and the MSD determined from elastic scans from a backscattering spectrometer.
This can be seen when measuring the scattering function $S(q,\omega)$ with different instruments (TOF, backscattering) over a large range of energy transfers. The Fourier transform $S(q,t)$ is displayed in Figure 7 as a function of temperature [36]. It is obvious that the combination of different experimental techniques is needed in order to cover a large enough time window of the intermediate scattering function. The indications from the NSE and elastic scan data from [35] are also corroborated since the decay of $S(q,t)$ in Figure 7 is not a single exponential function, but rather a distribution of relaxation times, which can be deduced from the data [37]. The proton diffusion is therefore not following the simple case of a single diffusion constant, but rather a distribution of constants due to different microscopic transport processes are present in the sample.

6.1 Conductivity and local proton transport

The local proton diffusion measured with quasielastic neutron scattering experiments or determined in computer simulations can be related to the macroscopically important proton conductivity with some simplifying arguments. We do this for the case of a PBI membrane of

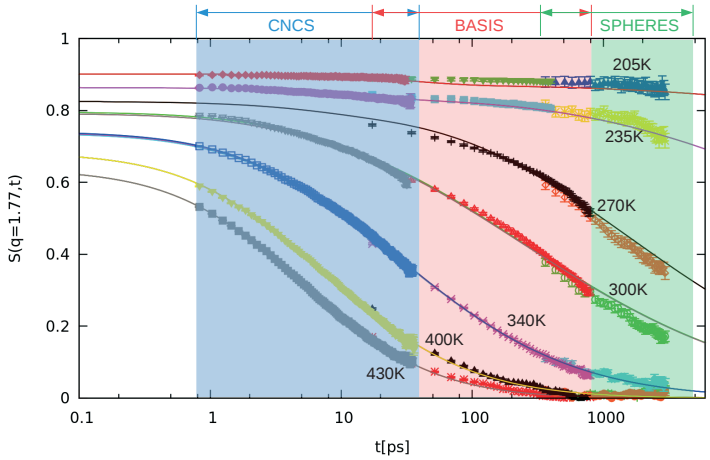


Fig. 7: *Intermediate scattering function obtained by Fourier transformation of the scattering function $S(q, \omega)$ from neutron scattering experiments (Time-Of-Flight, Backscattering).*

HT-PEFCs swollen with phosphoric acid (PA) [35]. The Nernst-Einstein equation links the proton diffusion with the macroscopic conductivity:

$$\sigma = N_V D_H e^2/(k_B T) \tag{14}$$

with the charge e and assuming a single diffusion constant D_H, with the temperature T, the Boltzmann constant k_B and the number of charges per unit volume, N_V. We estimate N_V (with 3 protons per PA molecule) to be $N_V = V/M \cdot \rho \cdot N_A \cdot 3 = 3 \times 10^{28}$ m^{-3}, where the bulk density of phosphoric acid of $\rho = 1.7$ g/cm^3 and a molar mass of $M = 98$ g/mol has been used [35]. Taking $D_H = 18 \times 10^{-11}$ m^2/s from NSE experiments [35] leads then to a conductivity of $\sigma = 24$ S/m, which is about half the value of the conductivity of pure phosphoric acid.

7 Summary

Fuel cells, in particular the HT-PEFCs which were in the focus of this article, are a complex multi component system. Its understanding on different length- and time scales is of importance for improving and further developing the technique. The combination of different experimental techniques helps largely in achieving this task. Computer simulations allow to study structural and dynamic properties which are inaccessible experimentally. Experiments on the other hand are required to verify the validity of simulations and to observe with intuitive models fundamental properties. Neutron scattering for example, which has the required sensitivity to light elements such as hydrogen, allows to measure correlation functions in space (structure of the membrane) and in time (diffusion of the proton). Transport processes on a molecular scale can be related in this way to macroscopic properties such as conductivity.

References

[1] K.-D. Kreuer (ed.), *Fuel Cells* (Springer, New York, 2013)

[2] K.A. Page, C.L. Soles, J. Runt (eds.), *Polymers for Energy Storage and Delivery: Polyelectrolytes for Batteries and Fuel Cells* (ACS Symposium Series 1096, ACS, Washington, DC, 2012)

[3] K. Schmidt-Rohr, Q. Chen, *Nature Materials* **2008** 7 75 – 83.

[4] D. E. Moilanen, D. Spry, M. Fayer, *Langmuir* **2008** *24* (8) 3690–3698.

[5] J.M. Serpico, S.G. Ehrenberg, J.J. Fontanella, X. Jiao, D. Perahia, K.A. McGrady, E.H. Sanders, G.E. Kellogg, G.E. Wnek, *Macromolecules* **2002** *35* 5916-5921.

[6] T. Ueki, M. Watanabe, *Macromolecules* **2008** *41* 3739–3749.

[7] J. Maranas, in K.A. Page, C.L. Soles, J. Runt (eds.), *Polymers for Energy Storage and Delivery: Polyelectrolytes for Batteries and Fuel Cells* (ACS Symposium Series 1096, ACS, Washington, DC, 2012)

[8] K. Sinha, J. Maranas *Macromolecules* **2014**, *47*, 2718–2726.

[9] P. Carlsson, R. Zorn, D. Andersson, B. Farago, W.S. Howells, J. Börjesson *J. Chem. Phys.* **2001**, *114*, 9645–9656.

[10] http://neutronsources.org/resources/educational-material.html

[11] J. S. Higgins, H. Benoît, Polymers and neutron scattering, Clarendon Press Oxford, 1994.

[12] J. Teixeira, *J. Appl. Cryst.* **1988** *21*, 781

[13] R. Zorn, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **2009** *603* (3) 439–445.

[14] S. Suarez, N. Kodiweera, P. Stallworth, S. Yu, S. Greenbaum, B. Benicewicz, *The Journal of Physical Chemistry B* **2012** *116* (41) 12545–12551.

[15] W. Maier, T. Arlt, K. Wippermann, C. Wannek, I. Manke, W. Lehnert, D. Stolten, *Journal of The Electrochemical Society* **2012** *159* (8) F398–F404.

[16] W. Maier, T. Arlt, C. Wannek, I. Manke, H. Riesemeier, P. Krüger, J. Scholta, W. Lehnert, J. Banhart, D. Stolten, *Electrochemistry Communications* **2010** *12* (10) 1436–1438.

[17] S. Pahari, C. K. Choudhury, P. R. Pandey, M. More, A. Venkatnathan, S. Roy, *The Journal of Physical Chemistry B* **2012** *116* (24) 7357–7366.

[18] L. Vilčiauskas, M. E. Tuckerman, G. Bester, S. J. Paddison, K.-D. Kreuer, *Nature Chemistry* **2012** *4* (6) 461–466.

[19] L. Vilčiauskas, M. E. Tuckerman, J. P. Melchior, G. Bester, K.-D. Kreuer, *Solid State Ionics* **2013** *252* (0) 34 – 39, sSPC-16 Conference Proceedings {SSI} Special Issue.

[20] B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.***2008** *4* 435447.

[21] G.A. Luduena, T.D. Kühne, D. Sebastiani, *Chemistry of Materials* **2011** *23* 1424–1429.

[22] K.-D. Kreuer, S. J. Paddison, E. Spohr, M. Schuster, *Chem. Rev.***2004** *104* 4637–4678.

[23] G. Dorenbos, K. Morohoshi, Energy Environ. Sci. **3**, 1326 (2010)

[24] R. Zorn in *Scattering Methods for Condensed Matter Research: Towards Novel Applications at Future Sources* (Schriften des Forschungszentrum Jülich, vol 33, Jülich, 2012)

[25] Kim, Sang-Hoon. *Journal of Korean physical society 46* (2): 474477. arXiv:cond-mat/0409763.

[26] I. Grillo, in *Soft Matter Characterization* (Springer, New York, 2008), 723-782

[27] O. Porod, in O. Glatter, O. Kratky (eds.), *Small Angle X-Ray Scattering* (Academic Press, London, 1982)

[28] Teubner M., and Strey R., *J.Chem. Phys.* **1987**, *87*,3195.

[29] L. Rubatat, A.L. Rollet, G. Gebel, O. Diart, *Macromolecules* **2002**, *35*,4050–4055.

[30] R. Devanathan, *Energy Env. Sci.* **2008**, *1*,101–119.

[31] , Th. Dippel, K.D. Kreuer, J.C. Lassegues, D. Rodriguez *Solid State Ionics* **1993**, *61*, 41–46.

[32] B. Frick, L. Vilčiauskas, P. P. Deen, S. Lyonnard, *Solid State Ionics* **2013**, *252*, 26–33.

[33] , K.A. Page, B.W. Rowe, K.A. Masser, A. Faraone *J. Polym. Sci. Part B: Polymer Physics* **2014**, *52*, 624–632.

[34] , J.-C. Perrin, S. Lyonnard, F. Volino *J. Phys. Chem. C* **2007**, *111*, 3393–3404.

[35] O. Holderer, O. Ivanova, B. Hopfenmüller, M. Zamponi, W. Maier, A. Majerus, W. Lehnert, M. Monkenbusch, R. Zorn, *Int. J. of Hydrogen Energy*,**2014** *39* 21657–21662

[36] B. Hopfenmüller, O. Ivanova, O. Holderer G. Ehlers, N. Jalarvo, M. Zamponi, W. Maier, A. Majerus, W. Lehnert, M. Monkenbusch, R. Zorn, *in preparation*

[37] F. Alvarez, A. Allegria, J. Colmenero, *J. Chem. Phys.* **1995**, *103*, 798–806.

G 4 Polymer architecture and function

E. van Ruymbeke
Bio and Soft Matter, IMCN
Université catholique de Louvain, Louvain-La-Neuve

Contents

1 Introduction 2

2 Fundamentals of tube models 2
2.1 Relaxation modulus 3
2.2 Rouse relaxation in entangled polymer melts 5
2.3 Disentanglement mechamisms 6

3 Linear viscoelasticity of complex molecular architectures 9

4 Combining rheology and tube models as a new characterization tool ... 11
4.1 Sensitivity of rheology to molecular artchitectures 11
4.2 Determining the composition of a polymer from its viscoealstic response (inverse modelling) 12
4.3 Combining rheology and statistical tools for determining the polymer composition 13

4 References 15

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

These last years, many experimental studies have shown that the viscoelastic response of a polymer melt strongly depends on its composition, which includes both dispersity in molecular weight and dispersity in architectures [1-4]. Understanding the relationship between composition and dynamics is thus of crucial importance in order to determine the role played by specific architectures and to design new polymeric systems with desired properties. Furthermore, this would allow us to use rheology as a new characterization tool, i.e. starting from the viscoelastic properties of a polymer melt in order to determine its composition [5].

In this direction, several coarse-grained models have been proposed, which describe the macromolecules at a mesoscopic scale, and which allow taking into account the motion of the different architectures present in the polymer [6-8]. Most of these models are based on the tube theory introduced by de Gennes [9] and by Doi and Edwards [10] in order to explain how an entangled polymer (either melt or highly concentrated solution) relaxes the anisotropy and related stress induced by a step strain. This model reduces the complex dynamics of inter-chain topological interactions to a "single-body" process, in which the molecular environment of an observed chain (usually called the "test" chain) is represented by a mean field called the "tube" in which the chain is confined.

In the following Sections, we first present the main relaxation mechanisms in tube theories for describing the relaxation of entangled linear chains. Then, in Section 2, the concept of hierarchical relaxation, which is used for describing the relaxation of complex branched polymers, is presented. In the last Section, we present how both rheological data and tube models can be combined in order to determine or validate the polymer composition.

2 Fundamentals of tube models

As illustrated in Figure 1, according to the tube theory, the test chain, which is unable to cross another molecule with which it is entangled, is behaving as if it was confined in a tube since only motions parallel to the curvilinear tube axis are possible, while lateral motions are limited to a characteristic distance, the tube diameter *a* [1-4].

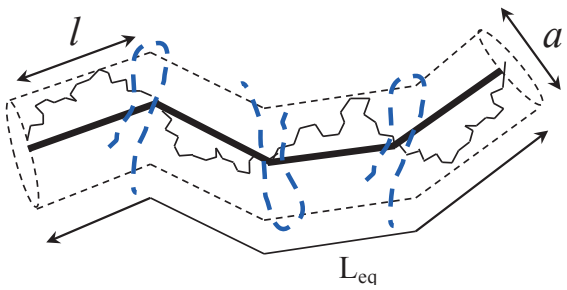


Fig. 1: *Schematic representation of a chain in the tube. Real chain: thin broken line; primitive path: thick line; Entanglements are represented by the dotted curves.*

Since the tube is an average object, its diameter has a constant value along the chain and for all chains. While it depends on the nature of the polymer and concentration, it is independent of chain architecture.

2.1 Relaxation modulus

The relaxation of a polymer after a small step strain is usually described by the relaxation modulus *G(t)*, the strain-independent ratio of the stress to the step-strain amplitude, which shows different regions in time (see Figure 2) [10,11].

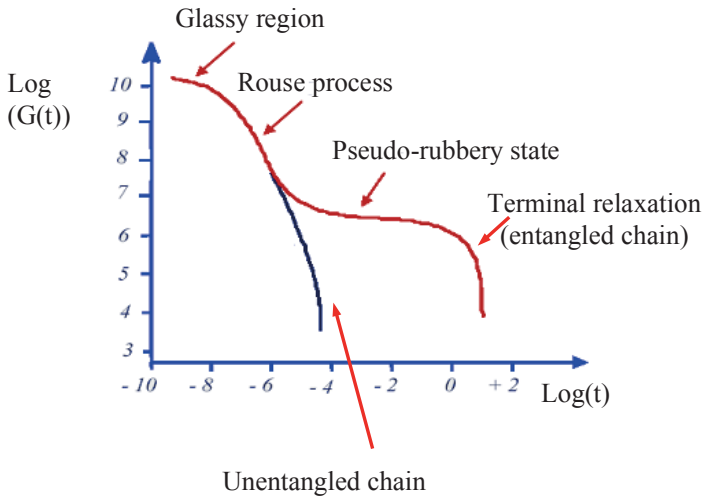


Fig. 2: *Different stages of the relaxation modulus of a polymer melt (see text for details).*

- After the glassy region**, the Rouse regime** describes the relaxation of the chains or subchains at times and length scale short enough to ignore mutual influences between the molecules. The chain motion is therefore described by the Rouse model [12], as described in Section 1.2.

- At longer times, the chains start interacting topologically and their motions are influenced by the environment: entanglements act as a temporary, physical network. At this stage, the tube representation becomes appropriate: embedded in the **temporary network**, the chains have reached their 'equilibrium' or **"rubbery" stage** and cannot relax further by the normal Rouse process, confined as they are by the surrounding tube. The main material parameters of the tube theory are defined at this stage: the plateau modulus G_N^0, the molecular weight between entanglements, M_e, which represents the molecular weight of the longer subchains relaxed at this stage, and the Rouse time of a segment between two entanglements τ_e, equal to τ_{Rouse}/Z^2, with τ_{Rouse} , the Rouse time of a test chain having Z entanglements. The material parameters G_N^0 and M_e are linked through the relation [1-4]:

$$G_N^0 = \frac{4}{5}\frac{\rho RT}{M_e}, \qquad (1)$$

where ρ is the density of the polymer and T the temperature. According to the precise definition of M_e, the pre-factor 4/5 is required or not in Equation 1. This has been clarified by Larson et al. [13].

Since it is assumed that a sub-chain between two entanglements is Gaussian, its end-to-end distance l is given as $l^2 = N_e b^2$, where b is the length of a Kuhn segment and N_e is the number of Khun segments between two entanglements [2]. It is useful to imagine that in the tube picture, the chain is coarse-grained at the scale of the segments between entanglements (Figure 1). They define the primitive path of the chain, having a length $L_{eq}(M)=Z.l=(M/M_e)l$. The primitive path can be considered as a (discretized) representation of the curvilinear axis of the tube [2].

Since we assume a Gaussian chain confined in a tube of diameter a, the most probable curvilinear path length L_{eq} is proportional to molecular weight through the relation:

$$R^2 = L_{eq}\, a = Zla, \tag{2}$$

where R is the quadratic end-to-end distance of the relaxed chain. On the other hand:

$$R^2 = Nb^2 = ZN_e b^2 = Zl^2, \tag{3}$$

where $N=ZN_e$ is the total number of Kuhn segments.

Hence: $a=l$.

- The second transition in Figure 2 corresponds to the **escape of the chain from the stressed entanglement network toward the relaxed equilibrium state**, via tube renewal. This includes different relaxation processes (reptation, contour length fluctuations, constraint release mechanisms) described in detail below.

Neglecting the glassy region, the relaxation modulus of an entangled polymer, $G(t)$, is thus described by two terms: the first takes into account the Rouse relaxation and the second describes the tube renewal process, which is proportional to the unrelaxed fraction, $F(t)$, of the polymer:

$$G(t) = G_{Rouse}(t) + G_N^0 . F(t). \tag{4}$$

Equation 4 is obviously not valid at very short times corresponding to the glassy region. Different models based on the generic tube theory have been proposed in order to correctly describe the relaxation function $F(t)$ of an entangled, well defined, polymer melt. While significant qualitative and quantitative differences are found between the various models, they all use the same basic ingredients to describe the relaxation of a chain: reptation, contour length fluctuations and constraint release mechanisms. These processes are introduced in Sections 1.3.

2.2 Rouse relaxation in entangled polymer melts

As shown in Figure 2, the high frequency (or equivalently short time) relaxation of an entangled polymer is described by a Rouse relaxation. A distinction is conveniently made between times shorter and longer than τ_e. At times shorter than τ_e, the relaxing segments, which are shorter than the distance between entanglements (itself equivalent to the tube diameter), do not feel the tube. At longer times, the segments feel the constraint imposed by the surrounding chains and therefore only the longitudinal Rouse modes along the tube are available [14]. This leads to the following expression for the Rouse modulus, including longitudinal modes:

$$G_{Rouse}(t) = G_e \sum_{p=Z}^{N} \frac{1}{Z} \exp\left(\frac{-2p^2 t}{\tau_{Rouse}(M)}\right) + \frac{G_e}{5} \cdot \sum_{p=1}^{Z-1} \frac{1}{Z} \exp\left(\frac{-p^2 t}{\tau_{Rouse}(M)}\right) \quad (5)$$

where $\tau_{Rouse}(M) = \tau_e \cdot Z^2$ is the Rouse time of a chain of molar mass M.

The first term accounts for the free Rouse relaxation of the segments shorter than the distance between entanglements (see Figure 3). The second term accounts for the longitudinal modes, confined by the tube to one dimension. The longitudinal modes are sometimes referred to as the "monomer redistribution process". Since the Doi and Edwards theory predicts the plateau modulus to be 4/5 of the rubber modulus because of the longitudinal motions along the tube, Likhtman et al. concluded that 1/5 of the stress stored in the tube after a step deformation is relaxed by longitudinal modes [14]. Therefore, the Rouse relaxation of an entangled chain can be expressed from its entanglement modulus, G_e, which is about 20% higher than the plateau modulus.

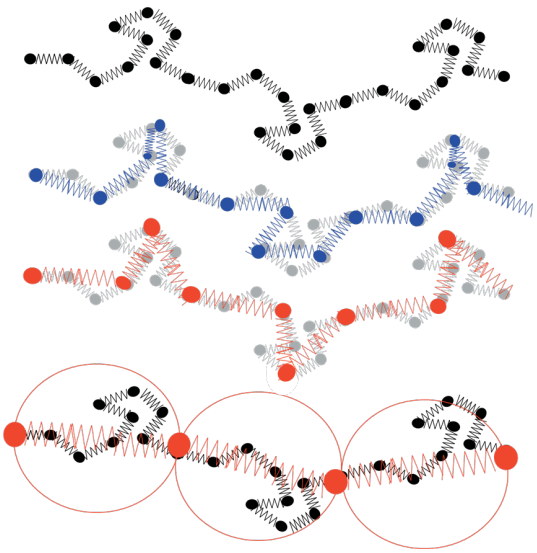


Fig. 3: *Illustration of the different mode of the Rouse relaxation of a polymer chains. First, fast modes take place, which involve the motion of 1 Kuhn segment. Then, slower modes including 2 Kuhn segments, etc. take place until the chain motion are prevented, due to the entanglements (see the last figure). This picture corresponds to the pseudo-equilibrium state (see Figure 2).*

2.3 Disentanglement mechamisms

Reptation

At times larger than the Rouse time of an entangled segment, τ_e, the tube picture is becoming active: Topological constraints restrict the motion of the chain laterally. Therefore, in order to relax its initial orientation, the chains need to find diffusion mechanisms which allow it escaping from this initial tube. In case of linear chains, the main relaxation process consist in the diffusion of the center of mass of the chain, along the curvilinear axis of the tube [4,9]. This process is defined by a one dimensional diffusion equation (Doi and Edwards, [10]). By back and forth motions along the primitive path (which are associated to a Brownian motion of the center of mass), the chain can move few by few outside the initial tube to create a new one. Reptation process is illustrated in Figure 4.

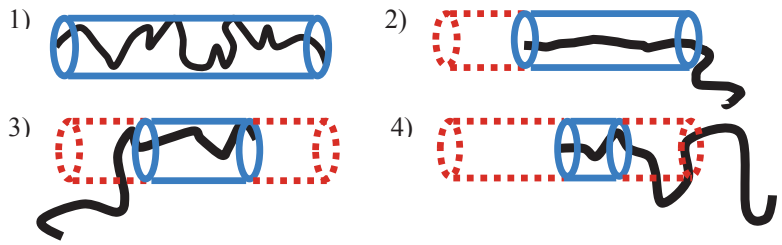


Fig. 4: *Cartoon of the reptation process. The initial tube is represented by blue lines. As soon as the chain diffuses along the tube axis, the memory of the initial orientation is partially lost.*

The chains will relax from the outer to the inner segments. According to reptation theory [9,10], the time needed to relax the entire chain by reptation, τ_d, is proportional to the cube of the molecular weight of the test chain:

$$\tau_d = \frac{{L_{eq}}^2}{\pi^2 D_c} \propto \frac{M^2}{M^{-1}} = M^3 , \tag{6}$$

where D_c, the curvilinear diffusion coefficient, is equal to $kT/N\zeta_0$, results from the summed drag of all the monomers along the chain and is thus proportional to the number of monomers, N, and the monomeric friction coefficient ζ_0. Expressing the primitive path length L_{eq} and the curvilinear diffusion coefficient in terms of tube model parameters gives the important relation [1-4]:

$$\tau_d(M) = \frac{\zeta_0 \; N_e^2 b^2}{\pi^2 kT} \left(\frac{M}{M_e} \right)^3 = 3 \; \tau_e Z^3 . \tag{7}$$

This M^3 dependence of the reptation time must be compared to the M^2 dependence of the Rouse time. The same M^3 scaling is predicted for the zero shear viscosity, η_o, since:

$$\eta_0 \propto G_N^0 . \tau_d , \tag{8}$$

and the plateau modulus is (within this picture) molecular-weight independent.

Contour length fluctuations and chain retraction

Because the experimental dependence observed for zero shear viscosity of entangled linear polymer melts follows a $M^{3.4}$ scaling rather the anticipated M^3 predicted by pure reptation, it is clear that additional relaxation mechanisms besides curvilinear diffusion of the center of mass are required to correctly describe the relaxation of linear chains [1-4]. Initially proposed by Doi [10] and later expanded by Milner and McLeish [15], **contour length fluctuations (CLF) mechanisms** is the second key ingredient to model the relaxation of polymeric architectures. Taking place at short times, this process takes into account the chain extremities are not constrained by the tube and therefore, can easily explore their surrounding environment. CLF is thus a Rouse relaxation of the chain ends, which does not require the motion of the center of mass. As shown in Figure 5, when the chain contracts within the tube and then stretches out again, the orientation of the ends of the initial tube is forgotten, and the stress associated with those portions is relaxed.

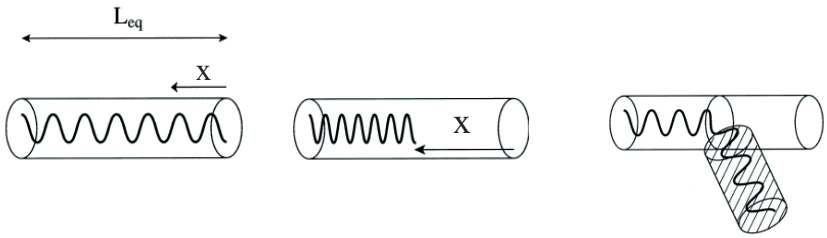


Fig. 5: *Illustration of the Contour Length Fluctuations process (taken from [4]).*

CLF will therefore speed up the overall relaxation of the polymer. The experimental 3.4 scaling for viscosity and reptation time of linear chains has been attributed to CLF by Doi [10], arguing that the fraction of tube relaxed by CLF should scale as $Z^{-1/2}$. The estimated time to relax a molecular segment localized at a position x along the chain (from 0 at the chain extremity to 1 in the middle of the chain) is determined as:

$$\tau_{early}(x) \approx \frac{9\pi^3}{16} \cdot \tau_e \cdot \frac{(x \cdot L_{eq,0}/2)^4}{a^4} \tag{9}$$

According to Doi and Edwards, the linear chain fraction relaxed by CLF does not have to relax by reptation, which speeds up this last process.

On the other hand, branched architectures, such as star-like molecules, are not able to relax by reptation. Therefore, while the extremities of the arms of the molecule will relax by CLF, the central molecular segments will relax by deep retraction within their tube. These deep retractions are called **chain retraction** [15,16]**.** They arise because the equilibrium length of a chain is only the most probable length, representing the most stable configuration. However, the real length of the chain fluctuates around this equilibrium length, allowing deeper molecular segments to relax. The retraction process is entropically unfavorable, and therefore characterized by long relaxation times, which increase exponentially with the molecular weight of the observed branch, or with the depth of the molecular segment along this branch:

$$\tau(x) = \tau_0(x) \ \exp\left(\frac{\nu x M}{M_e}\right), \tag{10}$$

Where ν is constant close to 3, x is the localization of the molecular segment along the branch (from 0 at the extremity to 1 at the branching point), M is the molecular weight of the branch and M_e is the molecular weight between two entanglements [1-4, 15-17].
Chain retraction is a typical first passage problem: it is not related to the most probable location of a chain end, but rather to the deepest tube segment reached by the fluctuating chain end.

Constraint release and tube dilation
Since the tube represents the topological constraint on a given chain from its molecular environment, the tube itself is able to move according to the motion of the surrounding chains. This tube motion, neglected in the Doi and Edwards model, is of particular importance for polydisperse systems. Tube motions are extremely complicated to describe in their generality and therefore only simplifications of the real situation are tractable. Tube motions chiefly affect the test chain in two different but related ways. First, "slow" tube motions allow large-scale lateral motions of internal segments, i.e. they induce a constraint release mechanism (CR) [4, 18], see Figure 6. Second, when motions of surrounding chains become fast at the observed time scale, the effect is equivalent to an increase of the "effective" tube diameter, which leads to an acceleration of the other relaxation processes (reptation, CLF). This mechanism is called "Dynamic Tube Dilation" (DTD) [19]. Several authors, using different approaches, have studied these two fundamental concepts.



Fig. 6: *Constraint release process (from [4]).*

According to the **Dynamic Tube Dilution** concept (DTD), which has been proposed by Marrucci [19], the "effective" tube of constraints around a chain widens as the relaxation proceeds. Indeed, the tube diameter is directly linked to the average distance that a chain can laterally cover without hitting topological constraints by the surroundings molecules. Were these obstacles fixed (i.e. a permanent network), the tube diameter would be a well-defined, time-independent quantity. However, because of the mobility of the surrounding chains, obstacles continuously disappear and reform, some of them more rapidly (near the chain end), others more slowly (far from the chain ends).

Therefore, the test chain will be able to move laterally and thus to explore the surroundings more and more with time. In other words, the tube diameter must be taken as an increasing

function of time during relaxation. This increase is calculated by assuming that the relaxed part of the polymer behaves like a solvent. Therefore, the effective tube diameter depends on the survival fraction of initial tube (i.e. the unrelaxed fraction of the polymer, considering reptation, CLF or retraction processes), $\Phi(t)$, as:

$$M_e(t) = \frac{M_e(0)}{\Phi(t)^{\alpha}} \tag{11}$$

$$L_{eq}(t) = L_{eq}(0).\left(\Phi(t)\right)^{\alpha/2} \tag{12}$$

$$a(t) = \frac{a(0)}{\left(\Phi(t)\right)^{\alpha/2}} \tag{13}$$

Note that the time dependent values are described as "effective" and are different from the true material parameters (except at time zero). Indeed, since the polymer density remains constant through time, the equilibrium values do not change.

From Equation 1 and Equation 11, we obtain:

$$G(t) = \frac{\rho RT}{M_e(t)^{\alpha}}\Phi(t) = G_N^0\Phi(t)^{1+\alpha} \tag{14}$$

The exponent α, called the dilution exponent takes a value between 1 and 1.3. Still today, there is no real consensus on its value [20].

In the DTD model, the relaxed part of the polymer is immediately taken as solvent. However, Struglinsky and Graessley [21] have shown that this assumption is not always. It works only if the different relaxation times are well separated on the time scale.

Ball and McLeish have extended the Dynamic Tube Dilation concept to chain retraction, in order to explain the relaxation of branched polymers [18]:

$$\frac{\partial \ln \tau_{fluct}\left(x_i,t\right)}{\partial x_i} = 3\cdot\left(\frac{M_i}{M_{e0}}\right)\cdot x_i\cdot\Phi\left(t\right)^{\alpha}, \tag{15}$$

This model was improved by Milner and McLeish [19] who solved the first passage problem to obtain more accurate expression of the initial retraction time and who proposed a transition equation to consider both early (non-activated) Contour Length Fluctuations and retraction process.

3 Linear viscoelasticity of complex molecular architectures

As for the linear chains, the linear viscoelastic response of complex architectures results from a combination of different relaxation mechanisms such as reptation, retraction and constraint release mechanisms. However, the motion of the polymer fraction which is trapped between two branching points requires, first, the motion of these branching points, which significantly slows down the relaxation of these inner segments. This leads to the important concept of **hierarchical motion**, according to which the outer generation of branches relax first, followed by the relaxation of the inner branches. This hierarchical relaxation is illustrated in Figure 7, which shows the storage and loss moduli of a polymer: while the outer branches relax at relatively high frequencies (around 100 rad/s), a second, low frequency plateau in the storage modulus is clearly observed (around 0.1 rad/s), which corresponds to the unrelaxed part of the chains, located between two branching points. Only after much longer time, the inner part of the molecules will also be able to move and relax, at the rhythm of the motion of the branching points.

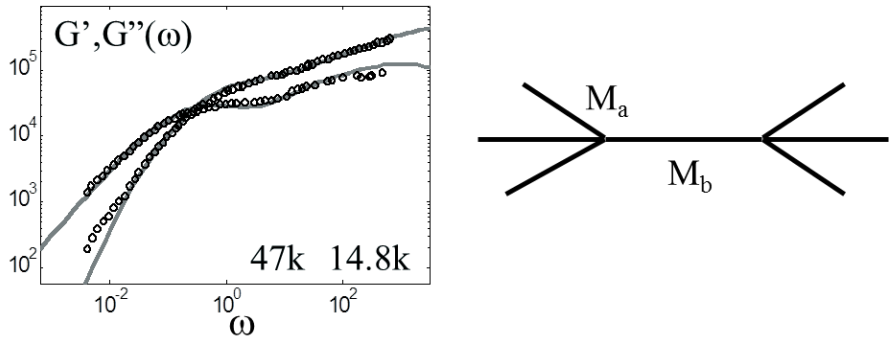


Fig. 7: *Storage and loss moduli of a polybutadiene sample with a pom-pom shape with the molar mass of an arm of 14.8 kg/mol and the molar mass of the inner backbone of 47 kg/mol. Since the inner part of the backbone is trapped between two branching points, its relaxation will take place at longer time, at the rhythm of the motion of the branching points. Symbols represent the experimental data (from ref. [22]), while the continuous curves represent the predictions based on a tube model [23].*

This hierarchy of relaxation processes in complex architectures has been observed with many complex model systems such as H-polymers, comb polymers [24] and Cayley tree polymers [25]. It must be noted that with polydisperse systems, this hierarchical process cannot be easily observed, due to the larger number of different relaxation times present in the polymer.

In order to quantify the relaxation time of the inner fraction of a branched molecule, extra friction coming from the branching point needs to be included in the reptation and retraction mechanisms. This has been proposed by several authors in the case of model polymer melts. Also, it has been generalized to polydisperse systems in refs. [26-28]. The idea behind these models is illustrated in Figure 8: first, the outer branches will relax. After its relaxation, each branch can be seen as an extra friction, of a certain importance, which is determined from the relaxation time of the branch.

As illustrated in Figure 7, accounting for this extra friction in the motion of the inner part of the molecules, a good agreement between theoretical and experimental viscoelastic curves has been found, which validated this approach.

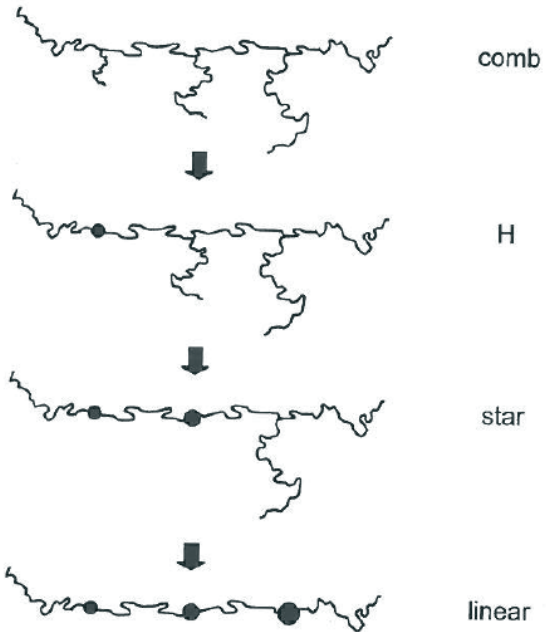


Fig. 8: *Illustration of the hierarchical relaxation of the branched macromolecule. After the deformation, only the branches can move by retraction. When a branch fully relaxes, it is replaced by a bead at the branching point, to take into account the frictional drag of the branch. Figure taken from ref. [1].*

4 Combining rheology and tube models as a new characterization tool

4.1 Sensitivity of rheology to molecular artchitectures

As explained in Sections 2 and 3, the **viscoelastic properties of a polymer melt strongly depend on its composition**. Given its high sensitivity, rheology is often used as an accurate tool to determine the architectures present in a complex polymeric material [5]. For example, it is used to detect low level of long-chain branching [29]. Indeed, comparing the experimental viscoelastic data to predictions based on linear chains, it is possible to estimate the level of branching chains in the sample. Furthermore, several studies with polyethylene samples inferred the presence of complex molecules from the thermo-rheological complexity of the material [30].

Thus, the viscoelastic curves of a model polymer include **the signatures of the branched architectures** present in the material, which can be used in order **to better characterize the sample composition**. In order to validate the consistency between linear viscoelastic properties and sample composition, models such as the tube models are very useful since they allow successfully predicting the linear rheology of polydisperse complex polymers, essentially without adjustable parameters. Thus, they allowed detecting low amount of side-

products in model polymers, which need to be taken into account in order to relate viscoelastic properties and sample composition, and despite the fact that these samples were considered as pure from SEC/LS analysis [5].

Furthermore, besides linear rheology, several recent works have shown the huge sensitivity of non-linear rheology to architecture and hence its potential to characterize such complex polymers. In particular, the strain hardening observed in extension can give us important information about the branching level and seniority of the molecules [31]. Quantitative models able to predict the non-linear behavior of polymer melts will certainly be further developed in the near future.

4.2 Determining the composition of a polymer from its viscoealstic response (inverse modelling)

While predicting the viscoelasticity of a polymer from its composition is well-defined, determining the composition of a polymer from its viscoelastic curves represents an ill-posed problem[1]: rheology alone does not suffice to determine the composition of a sample, irrespectively of the viscoelastic model used. In order to resolve this issue and be able to extract molecular structure from the viscoelastic data requires describing the whole sample composition based only on few parameters in order to minimize the number of possible solutions (and hence, uncertainty).

For example, as illustrated in Figure 9, determining molar mass distribution (MMD) of a sample from its viscoelastic properties has been applied to linear polymers. Indeed, knowing *a priori* that a sample is linear, it was shown that the inverse problem could be solved by approximating the unknown MMD by one or two log-normal distributions or Generalized exponential (GEX) functions [31]. In this way, the number of unknown parameters was low enough (two parameters by log-normal and three by GEX function) to overcome the ill-posed properties of the inverse problem. The parameters related to the MMD were subsequently determined by means of a least-square minimization.

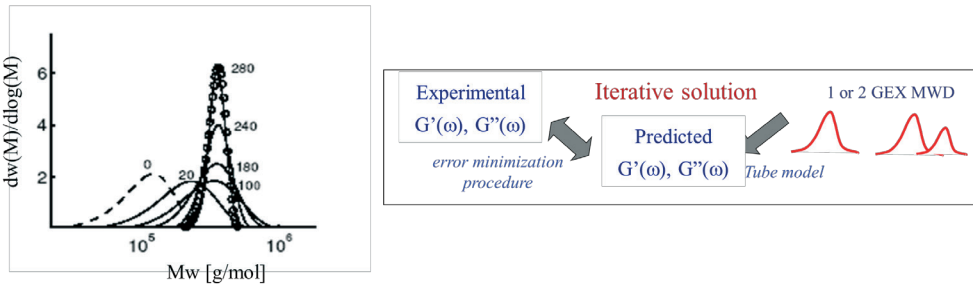


Fig. 9: *Illustration of solving the inverse problem, starting from the rheological data to determining the MMD of a monodisperse linear polystyrene sample. In the left figure, the symbols represent the experimental data, while the continuous curves show the theoretical MMD obtained after a certain number of iterations. In this specific case, after 280 iterations, the model converges toward the experimental MMD and stops evolving.*

Despite these encouraging developments, in most cases even model star polymers are already too complex to allow transforming their viscoelastic response to MMD. Therefore, these mesoscopic models are rather used to validate a specific composition, proposed for example from the analysis of its MMD, or to further investigate the influence of specific architectures on the viscoelasticity of the sample.

4.3 Combining rheology and statistical tools for determining the polymer composition

While the MMD of a branched polymer can be determined experimentally, its interpretation in terms of different architectures contributing to the signal is far from trivial, since several architectures could lead to virtually the same signal. Hence, associating MMD to architectures remains a formidable challenge. In order to reduce the ill-posed character of this exercise, it is necessary to exploit any information available from the synthesis protocol.

For example, due to their nature or their synthesis protocol, some polydisperse branched polymers have a composition which follows well-defined statistical rules [5]. In this case, it has been shown that despite the large number of different macromolecules, the MMD measured by SEC/LS could be decomposed into different architectures without facing the ill-posed problem of model polymers. Decomposition of the SEC/LS data is indeed possible if the whole set of molecules could be determined from a few parameters like branching density and number-average molar mass.

A first example of such systems are the polydisperse branched samples which have been obtained by coupling different polydisperse linear parents. In such a case, one can strongly reduce the degree of freedom and thus the ill-posed properties for determining the possible architectures present in the polymer.
This has been illustrated in ref. [32], where a blend of linear and star chains were obtained by adding coupling agent in a linear polydisperse parent. In that case, only combinations of two (to create double linear), three (to create 3-arm stars) or four linear parents (to create 4-arm stars) were, chemically, possible (see Figure 10).

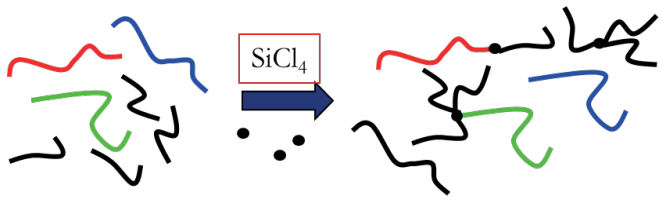


Fig. 10: *Statistical composition of a polymer blend obtained from a polydisperse linear polymer in which coupling agents have been added.*

The corresponding MMDs of these different species could be easily determined from the MMDs of the linear (polydisperse) parents. In this way, only three unknown parameters are needed to describe the statistical MMD of the whole sample: the proportions of linear, double linear and 3-arm star polymers (with the proportion of 4-arm star polymers being the remaining fraction of the polymer). Therefore, this inverse problem could be solved by using a best-fit procedure. This is illustrated in Figure 11.

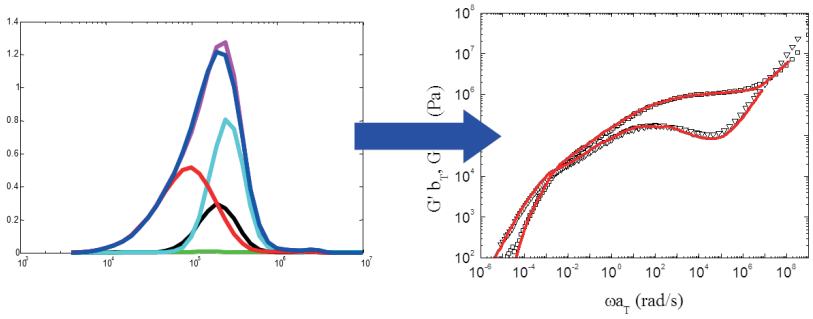


Fig. 11: *(left:) Decomposition of the experimental MMD (blue curves) into linear, double linear, 3arm stars and 4arm stars polybutadiene polymers. The sum of these different distributions is shown by the magenta curve. (right:) Comparison between experimental storage and loss moduli (o) and the moduli predicted based on the statistical composition.*

Another example of systems for which the ill-posed properties of the inverse problem can be reduced are poly-condensate samples [5]. Indeed, for these samples, the different architectures present in the polymer can be statistically determined based on a Monte Carlo simulation. In such a case, if the different chemical components are known, the only unknown needed to fit the experimental SEC/LS curve is the conversion level. Interpreting the SEC/LS curves becomes then possible, and the comparing the experimental viscoelastic data to the predicted curves based on this statistical distribution allows, again, to validate the approach. This has been illustrated, for example, in ref. [33].

4 References

[1] J.M. Dealy and R. G. Larson, Structure and rheology of molten polymers (Hanser Publishers, Munich, 2006)

[2] Rubinstein, M.; Colby, R. H. *Polymer Physics*; (Oxford University Press: New York, 2003)

[3] T. C. B. McLeish, Adv. Phys. 51, 1379 (2002)

[4] H. Watanabe, Prog. Polym. Sci. 24, 1253 (1999)

[5] E van Ruymbeke, H Lee, T Chang, A Nikopoulou, N Hadjichristidis, F Snijkers, D Vlassopoulos, Soft Matter **10**, 4762 (2014)

[6] Z. W. Wang, X. Chen and R. G. Larson, J. Rheol. 54, 223 (2010)

[7] Y. Masubuchi, G. Ianniruberto, F. Greco and G. Marrucci, J. Chem. Phys. 119, 6925 (2003)

[8] A. E. Likhtman, S. K. Sukumaran and J. Ramirez, Macromolecules 40, 6748 (2007)

[9] P. G. de Gennes, J. Chem. Phys. 55, 572 (1971)

[10] M. Doi and S. F. Edwards, The Theory of Polymer Dynamics (Oxford, 1986)

[11] van Ruymbeke, E.; Liu, C. Y.; Bailly, C., Rheology Reviews; Binding, D. M., Hudson, N. E., Keunings, R., (The British Society of Rheology, **2007**)

[12] Rouse PE, "A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers", J. Chem. Phys. **21** 1272 (1953)

[13] R. G. Larson, T. Sridhar, L. G. Leal, G. H. McKinley, A. E. Likhtman and T. C. B. McLeish, J. Rheol. 47, 809 (2003)

[14] Likhtman, A.E.; McLeish, T.C.B.; Macromolecules 35, 6332 (2002)

[15] S. T. Milner and T. C. B. McLeish, Macromolecules 30, 2159 (1997)

[16] S. T. Milner and T. C. B. McLeish, Macromolecules 31, 7479 (1998)

[17] Shchetnikava, V.; Slot, J. J. M.; van Ruymbeke, E.; Macromolecules, accepted (2014)

[18] des Cloizeaux, J., J. Europhys. Lett. **5**, 437 (1988)

[19] G. Marrucci, J. Polym. Sci., Polym. Phys. Ed. 23, 159 (1985)

[20] van Ruymbeke, E.; Masubuchi, Y.; Watanabe, H.; Macromolecules 45, 2085 (2012)

[21] M. J. Struglinski and W. W. Graessley, Macromolecules 18, 2630 (1985)

[22] L. A. Archer and S. K. Varshney, Macromolecules 31, 6348 (1998)

[23] E. van Ruymbeke, C. Bailly, R. Keunings, D. Vlassopoulos, Macromolecules 39, 6248 (2006)

[24] M. Kapnistos, D. Vlassopoulos, J. Roovers and L. G. Leal, Macromolecules 38, 7852 (2005)

[25] E. van Ruymbeke, K. Orfanou, M. Kapnistos, H. Iatrou, M. Pitsikalis, N. Hadjichristidis, D. J. Lhose and D. Vlassopoulos, Macromolecules 40, 5941 (2007)

[26] R. G. Larson, Macromolecules 34, 4556 (2001)

[27] C. Das, N. J. Inkson, D. J. Read, M. A. Kelmanson and T. C. B. McLeish, J. Rheol. 50, 207 (2006)

[28] E. van Ruymbeke, R. Keunings and C. Bailly, J. Non- Newtonian Fluid Mech. 128, 7 (2005)

[29] E van Ruymbeke, V Stephenne, D Daoust, P Godard, R Keunings, C. Bailly, Journal of rheology 49, 1503 (2006)

[30] P. Wood-Adams and S. Costeux, Macromolecules 34, 6281 (2001)

[31] E. van Ruymbeke, R. Keunings and C. Bailly, J. Non-Newtonian Fluid Mech. 105, 153

(2002)
[32] E. van Ruymbeke, S. Coppola, L. Balacca, S. Righi and D. Vlassopoulos, J. Rheol., 2010, 54, 507 (2010)
[33] E. van Ruymbeke, J. J. M. Slot, M. Kapnistos and P. A. M. Steeman, Soft Matter 9, 6921 (2013)

G 5 Micro- & Nanotechnologies for Neuroscience

A. Offenhäusser, J. Albers, S. Ullmann
Institute of Complex Systems, ICS-8, Peter Grünberg Institut, PGI-8
Forschungszentrum Jülich GmbH

Contents

1 Introduction 2

2 Nanofabrication 2
2.1 Lithography 3
2.2 Etching 3
2.3 Deposition 5

3 Interfacing Electronic Devices with Neuronal Cells 12
3.1 Working principles 12
3.2 Planar Microelectrode Arrays 14
3.3 3D Electrodes 17
3.4 Fieldeffect Transistor Arrays 22
3.5 High-Density CMOS-type MEA arrays 24

4 Patterning neuronal networks on electrode arrays 25
4.1 Controlling adhesion and outgrowth of individual neurons 26
4.2 Controlling connectivity of neuronal populations 26

References 30

Lecture Notes of the 46th IFF Spring School "Functional Soft Matter" (Forschungszentrum Jülich, 2015). All rights reserved

1 Introduction

The human brain is an organ of vast complexity and despite ever increasing effort, scientists are still far from understanding how this network of several billion neurons accomplishes functional interaction and signalling or to find reliable methods to assist this intricate machinery in places where correct signalling fails. Due to its complexity, *in vivo* studies are limited to macro-scale investigations such as electroencephalography (EEG), magnetic resonance imaging (MRI) and positron emission tomography (PET). While these methods allow for the ascertainment of organ function and the localization of its failures, it is impossible to characterize single cell function and the mechanisms of network communication from these studies. Therefore, our group focuses on the investigation of small-scale, two-dimensional neuronal networks (up to 1000 cells/mm^2), their interaction and possible means to influence their behaviour. We aim to establish methods for the manipulation of bionetwork development and establishment of connectivity patterns via microcontact printing (μCP) and novel microfluidic systems. Within this area of expertise, we engineer shapes and then deposit corresponding protein patterns via μCP that facilitate a preferential outgrowth of neurites in a predetermined direction, thereby controlling the directionality of signal propagation within the network. Furthermore, it is our goal to fabricate chip-based sensors that enable an efficient cell-chip coupling towards precise recording of cellular signals. We want to facilitate a better understanding of the cell-cell communication and provide the ability to stimulate network communication by the transmission of electrical signals to the constituting neurons. Within this framework, we have developed a variety of microelectrode array (MEA) designs that enable non-invasive, parallel, multi-site recording of action potentials from primary neurons and cardiomyocyte-like HL-1 cell line. We have modified our standard planar 64 electrode MEA design with different geometries ranging from nanometer-sized cavities that allow for cellular protrusion into the sensor to mushroom-shaped 3D electrodes. Furthermore, we investigate various field-effect transistor (FET) designs with gate materials ranging from silicon nanowires to graphene. In order to realize the development of our devices, we employ both traditional cleanroom technology as well as modern 3D printing procedures. Ultimately, the combination of our different approaches – microcontact printing, microfluidics and multi-electrode arrays – yields novel devices for the study of network development and communication.

This Chapter is meant to give an overview over the commonly used techniques, processes and approaches in the area of multielectrode arrays for the study of cellular communication. Section 1 describes the techniques of relevance for our nanofabrication procedure, Section 2 various MEAs designs and their respective advantages and Section 3 deals with the formation of specific connectivity patterns of neuronal cultures on MEA devices.

2 Nanofabrication

Nanotechnology focusses on the design, synthesis, and characterization, of multi-functional materials and devices that have a functional organization in at least one dimension on the nanometer scale, ranging from a few to about 100 nanometers. Applications of nanotechnology to biology and medicine are varied and continue to advance as the engineering continues to improve. In general, the preparation techniques in nanotechnology can be categorized as bottom-up or top-down, or a combination of both. Top-down

approaches start with a bulk material, in which a detailed structure is transferred. Methods such as lithography and etching techniques fall into this category. Bottom-up approaches start with a defined structure or a pre-conditioned material that undergo specific chemical or physical processes that produce higher-ordered structures. Methods such as self-assembly, templating and scaffolding fall into this category. In the following paragraphs concepts and examples will be described.

2.1 Lithography

Lithography is one of the key processes used in microfabrication and is used to pattern thin films or the bulk of a substrate. It uses light or electrons to transfer a predefined pattern to a light- or electron-sensitive material (resist) on the substrate (see Fig. 1). The exposure is followed by a series of chemical or physical process steps which either etches the pattern into, or deposits new material in the desired pattern onto, the material underneath the photo resist [1]. Exposure systems in photolithography typically produce an image on the wafer using a photomask. Exposure systems either work in contact or proximity mode or as projection systems called stepper. Current state-of-the-art photolithography tools use deep ultraviolet (DUV) light sources, which allow minimum feature sizes down to 50 nm and below. In contrast electron-beam lithography is a maskless method and directly writes patterns with high resolution (sub-10 nm) but low throughput.

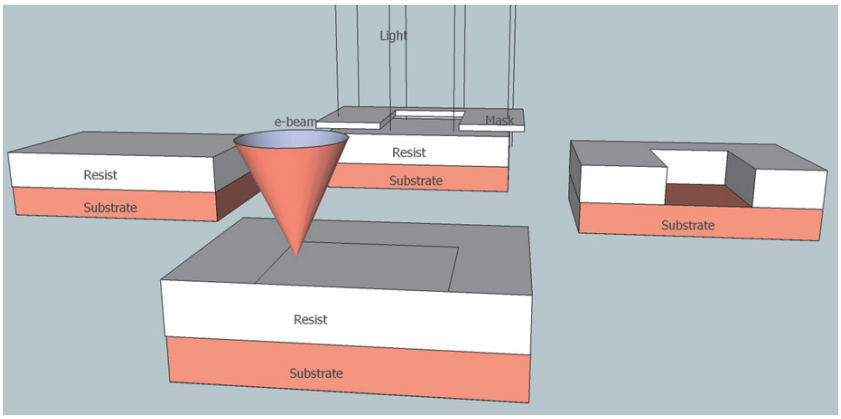


Fig. 1: *Optical and electron-beam lithography: the flat substrate is spincoated with photo or electron sensitive resist (left). Resists consist of macromolecules that are modified upon exposure to light (back) or high-energy electrons (front), resulting in a changed solubility. A desired sample pattern can thus be transferred to the resist by writing the pattern and subsequent wet chemical development (right). The resulting patterned resist layer serves as a mask for the following steps.*

2.2 Etching

The next step in making a structure on the wafer after lithography is called etching, by which layers of material are removed from the surface of a wafer by chemical or physical reactions. Usually, part of the wafer is protected from the etchant by a "masking" material which resists this reactions. In the simplest case, the resist used before in lithography can be used as masking material but normally a more resistant masking layer is required. The etching process

results in an undercut of the masking layer (see Fig. 2). The ratio between the width and the depth of a structure is called aspect ratio. Etchants are called isotropic when large undercuts are produced, resulting in small aspect ratio. Modern fabrication processes greatly prefer anisotropic etching, because they can produce high aspect ratio structures. Etching processes can be categorized in dry (gas phase) and wet (liquid phase) etching.

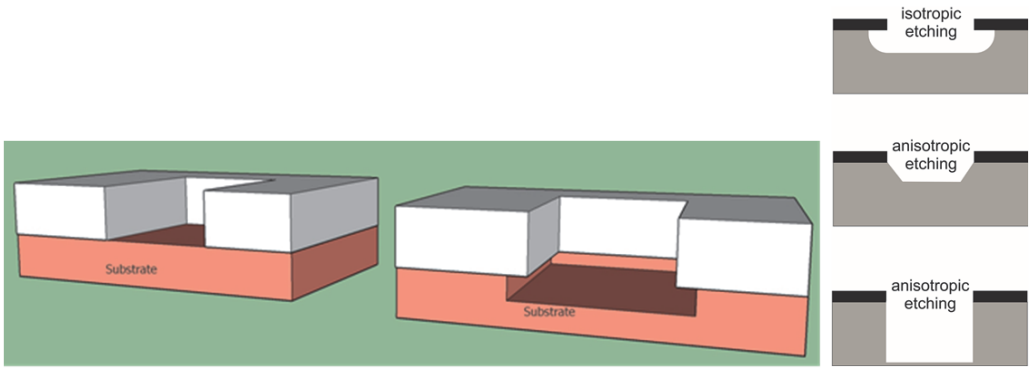


Fig. 2: *After lithography the pattern is transferred from the resist to the substrate by an etching step. Wet etching (chemical etching) and dry etching (plasma or ion etching, thus, physical etching) are both commonly used methods in the micro- and nanofabrication nowadays. Dry etching is preferred over the wet etching when patterns with finer geometries are required.*

Wet etching Wet etchants are usually isotropic (see Fig. 2, right), leading to a large undercut when etching thick films. Wet etching is normally a more simple technology as it requires a container with a liquid solution that will dissolve the material in question. A commonly used etchant for silicon dioxide over a silicon substrate is buffered hydrofluoric acid (BHF). The main challenge is to find a mask that will not dissolve in the etchant or at least etches much slower than the material to be patterned and to find a large etch rate difference between the material to be patterned and the material at which the etching process should stop.
Some wet etchants react with crystalline materials at very different rates depending upon which crystal face is exposed (see Fig. 2, right). In single-crystal materials (e.g. silicon wafers), this effect can allow very high anisotropy and high aspect ratio. Several anisotropic wet etchants are available for silicon. For instance, potassium hydroxide (KOH) has an etch rate selectivity 400 times higher in <100> crystal directions than in <111> directions. Tetramethylammonium hydroxide (TMAH) presents an alternative, with a 37 times selectivity between <100> and <111> planes in silicon.

Dry etching A disadvantage of wet etching is the undercut caused by the isotropy of the etching process. As an alternative to wet etching, plasma or dry etching is used which can create a more anisotropic etching, critical for high-aspect ratio pattern transfer. The plasma is generated under low pressure (vacuum) by an electromagnetic field. Energetic free radicals and high-energy ions are produced that attack the wafer surface and react with it.
Plasma etching can be isotropic, i.e., exhibiting a lateral undercut rate on a patterned surface approximately the same as its downward etch rate, or can be anisotropic, i.e., exhibiting a smaller lateral undercut rate than its downward etch rate. Such anisotropy is maximized in deep reactive ion etching.

Reactive-ion etching (RIE) is the most prominent type of dry etching. RIE uses chemically reactive plasma to remove material deposited on wafers. A typical (parallel plate) RIE system consists of a vacuum chamber, with a wafer placed on the bottom electrode (see Fig. 3). The amount (gas pressure) and composition of gas in the chamber heavily influences the etching process: e.g. fluoride based gas is commonly used for etching silicon.

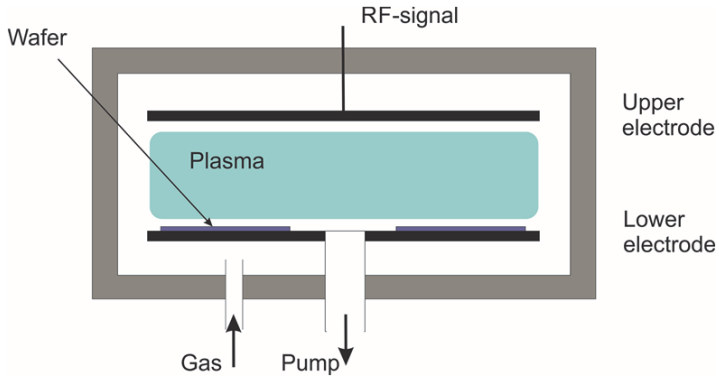


Fig. 3: *Sketch of a common RIE setup: A RIE consists of two electrodes that create an electric field meant to accelerate ions toward the surface of the samples.*

Other types of RIE systems exist, including inductively coupled plasma (ICP) RIE. In this type of system, the plasma is generated with an RF powered magnetic field. Very high plasma densities can be achieved, though etch profiles tend to be more isotropic.

A combination of parallel plate and inductively coupled plasma RIE is possible. In this system, the ICP is employed as a high density source of ions which increases the etch rate, whereas a separate RF bias is applied to the substrate (silicon wafer) to create directional electric fields near the substrate to achieve more anisotropic etch profiles.

To achieve near vertical structures a so-called **Bosch process** can be applied. The process alternates rapidly between two modes, a standard plasma etch and a deposition of an inert passivation layer. During the etching phase, the directional ions that bombard the substrate attack the passivation layer at the bottom of the trench (but not along the sides). They collide with it and sputter it off, exposing the substrate to the chemical etchant.

2.3 Deposition

The deposition of defined layers of metals or isolators is an important step for the majority of micro- and nanofabrication processes to define feedlines, bondpads and electrode areas and to prepare isolation layers to selectively enable contact to the environment in defined positions. The following paragraphs will describe common layer deposition techniques.

Metals

Physical vapor deposition (PVD) is the direct deposition of atoms from source to a target and is most commonly used for the formation of metallic thin layers. PVD can either be performed via thermal evaporation or sputtering. During thermal evaporation, the source material is heated inside a crucible either via resistive or e-beam heating in a low-pressure

environment, resulting in the evaporation and subsequent deposition on the substrate via condensation (Fig. 4a). This process exhibits very good directionality, generally resulting in poor edge coverage (Fig. 4b) and substrate orientation dependent deposition (Fig. 4c) [2]. This directionality is advantageous for lift-off processes where poor edge coverage enables the dissolution of the photoresist and results in clean edges of the patterned metal layer. For processes where a more homogeneous deposition with improved edge coverage is essential, moving sample carriers can be utilized to improve the layer characteristics.

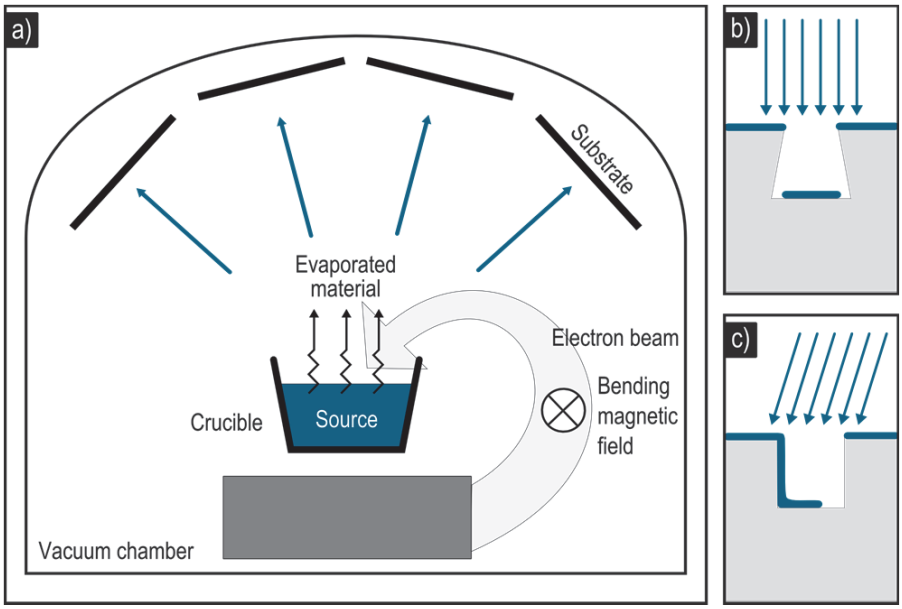


Fig. 4: *Schematic representation of physical vapor deposition via e-beam evaporation. The substance to be deposited (source) is heated inside a crucible, evaporates and condenses on the substrate (a). This process exhibits a distinct directionality, resulting in a poor edge-coverage ideal for lift-off processes (b) and the possibility of influencing the metal layer by changes to the substrate orientation (c) (schemes according to [2]).*

Sputtering is an alternative to thermal evaporation. In this case, the material to be deposited is bombarded with high-energy inert ions, e.g. argon, resulting in the ejection of atoms from the material towards the target. While this process results in better edge coverage, it has the disadvantage of higher substrate temperature as compared to thermal evaporation, which is problematic in case of the deposition of metallic layers on top of polymeric photoresists. Since many photoresists exhibit a low glass transition temperature T_g above which they become soft resulting in distortions, the substrate temperature is an important parameter for the deposition of metal thin films [3].

PVD machines are usually equipped with multiple source materials, which is a great advantage since it enables the deposition of stacks from different materials within the course of one processing step. This is especially important for noble metals such as gold and platinum which generally show poor adhesion on the substrates commonly employed during microfabrication. In this case, the addition of an interstitial layer of either chromium or titanium greatly improves the adhesion and thereby device stability.

A disadvantage of PVD processes is the limited deposition thickness achievable using these methods. Due to stress within the formed layers which can lead to the formation of cracks, layers thicker than 2µm are not commonly prepared using this technique [2]. Furthermore, since the source material condenses on substrate and reactor wall alike, this approach requires comparatively large amounts of material.

Galvanization or electroplating is another form of metal deposition. This method relies on the reduction of metal salts in aqueous solution and, in contrast to PVD, is also suitable for thicker deposits [2]. Fig. 5a depicts the principle of electroplating. At the cathode, positively charged metal ions are reduced to the elemental metal. The electrons needed for this process are either supplied via the oxidation of the anode if a base metal such as copper is employed

$$Cu \rightarrow Cu^{2+} + 2\ e^{-} \quad (1)$$

or via the oxidation of water

$$2\ H_2O \rightarrow 4\ H^{+} + O_2 + 4\ e^{-} \quad (2)$$

if an inert material is used as anode. The introduction of an electrically conductive layer (seed layer) underneath the photoresist during lithography based microfabrication enables the utilization of this process for the formation of structured metal areas (Fig. 5b). In areas, which are not covered by photoresist, the conductive layer, which functions as cathode, is in contact to the metal ions in solution. Reduction of said metal ions results in the formation of elementary metal deposits. The thickness of the deposit is controlled via the applied voltage and deposition time. Subsequent seed layer stripping produces free-standing metal structures.

One metal that is often used for electronic applications is gold since it exhibits good electrical and thermal conductivity, wear resistance and inertness, which prevents the formation of an insulating oxide layer [4]. Many gold electroplating formulations are commercially available, mostly based either on gold sulfite, thiosulfate or gold cyanide complexes. Due to their superior stability, gold cyanide complexes, such as potassium dicyanoaurate(I) ($K[Au(CN)_2]$) which exhibits a potential of -0.61V against the standard hydrogen electrode [4], are most commonly used. To prevent contamination, inert electrodes are utilized for the electrochemical deposition of gold with the electrons needed being supplied via the oxidation of water.

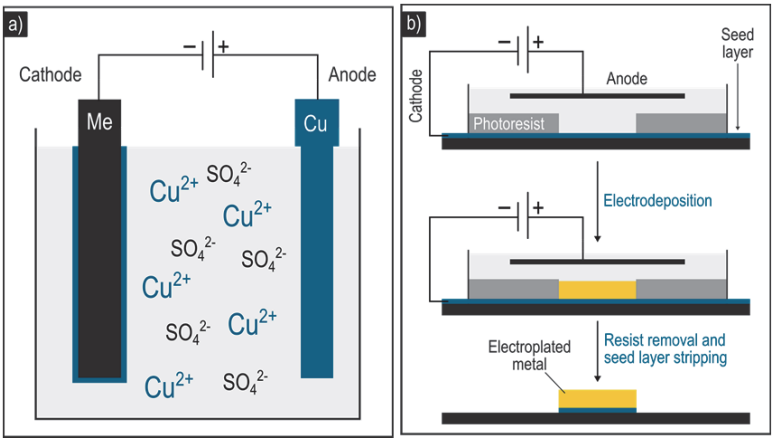


Fig. 5: *a) Principle of galvanization. Positively charged metal ions from solution are reduced at the cathode to form a metal layer. If an easily oxidized metal is used as anode, the metal ions in solution are constantly replenished, in which case, the anode is termed "sacrificial". If inert anodes such as platinum are used, electrons can be supplied by the oxidation of water. b) Application of electroplating in combination with microfabrication for the formation of metal patterns. A conductive layer (seed layer) underneath the photoresist enables the formation of an electrical contact and the electrochemical deposition of metals in areas which are not covered by photoresist. Subsequent photoresist removal and seed layer stripping yields free-standing metal structures.*

A clear advantage of electroplating over PVD is the ability to form thick deposits of several tens of µm. Furthermore, since the deposition only occurs "inside the pattern", i.e. in areas not covered by photoresist, it is more economical with no material wasted to cover the photoresist and be removed during lift-off or covering the reactor walls as is the case for PVD processes.

$$\begin{aligned} &\text{Cathode:} && 4\ K[Au(CN)_2] + 4\ e^- \rightarrow 4\ Au + 4\ KCN + 4\ CN^- \\ &\text{Anode:} && 2\ H_2O \rightarrow 4\ H^+ + O_2 + 4\ e^- \\ &\text{Full cell:} && 4\ K[Au(CN)_2] + 2\ H_2O \rightarrow 4\ Au + 4\ HCN + 4\ KCN + O_2 \end{aligned} \qquad (3)$$

Isolators

Thermal oxidation is a method that enables the formation of a thin layer of oxide and is most commonly used on silicon in order to prepare isolating layers of silicon oxide. While silicon naturally establishes an oxide layer even at room temperature, the reaction comes to a stop once the surface is fully covered and thus passivated, yielding an oxide layer of only around 2nm in thickness [5]. In order to achieve thicker layers, the reaction is performed in an oxidation furnace at high temperatures of around 900°C to 1200°C (Fig. 6a). Since the reaction is not a deposition in the classical sense but rather occurs at the silicon/silicon oxide interface under consumption of the silicon substrate, the solid state diffusion of the reactants through the established oxide layer exhibits a significant influence on the reaction kinetics, resulting in a transition from a surface-reaction limited case for thin oxides to the limitation by the diffusion of the oxidizing species to the reaction site for thicker oxides [5]. Due to the lower density of the established oxide as compared to the silicon substrate, the SiO_2 both grows "in" and "out" from the original Si surface (Fig. 6b,c). The original layer thickness of the silicon consumed during the reaction (X_{Si}) is approximately 0.46 times that of the resulting oxide (X_{Ox}) [6], every 1µm of silicon thus yields 2.17µm of SiO_2 after oxidation.

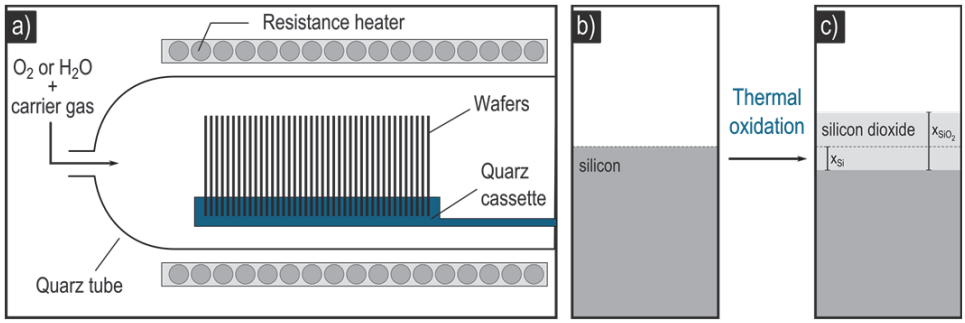


Fig. 6: *Thermal oxidation of silicon. The reaction is performed in a quartz tube, resistance heaters supply the required thermal energy (a). Multiple wafers can be inserted into the furnace and oxidized in parallel (scheme according to [2]). Depending on the oxidation mechanism, either plain oxygen or water vapor in a suitable carrier gas are introduced into the device, resulting in the reaction of silicon (b) to an insulating silicon oxide layer. Due to the consumption of silicon at the Si/SiO_2 interface and the lower density of the oxide, it can be found both above and below the line of the original silicon surface (c).*

Thermal oxidations of silicon can either be performed in the presence of plain oxygen (*dry oxidation*) [2]

$$Si_{(s)} + O_{2(g)} \rightarrow SiO_{2(s)} \tag{4}$$

or water (*wet oxidation*)

$$Si_{(s)} + 2H_2O_{(g)} \rightarrow SiO_{2(s)} + 2H_{2(g)} \tag{5}$$

The reaction rate is significantly smaller for dry oxidations as compared to wet oxidations. Nevertheless, the oxidation with oxygen exhibits certain advantages since it results in silicon oxide layers of superior quality due to an increased density, fewer defects and corresponding higher dielectric strength. However, since the process in extremely slow and thus time consuming, dry oxidations are only performed for thin oxide layers. Alternating dry-wet-dry cycles enable the combination of the advantages from both processes.

Chemical vapor deposition (CVD) enables the formation of various thin films by either reaction or decomposition of volatile precursors. Depending on the employed process parameters, nature of chose precursors or the method of supplying the energy needed for the precursors to either react or decompose to the intended product, the literature distinguishes between various different types of CVD processes such as low-pressure CVD (LPCVD), metal-organic CVD (MOCVD) or plasma-enhanced CVD (PECVD) [7]. CVD enables the deposition of various silicon and carbon containing compounds such as silicon oxide, silicon nitride, silicon carbide, carbon nanotubes and synthetic diamond. Furthermore, it is possible to deposit polysilicon by reduction of silanes [6].

$$SiH_4(g) \rightarrow Si(s) + 2H_2(g) \tag{6}$$

This LPCVD reaction is important both for the fabrication of complementary metal oxide semiconductor (CMOS) integrated circuits. However, since this reaction has to be performed

at temperatures of approximately 600°C and as high temperatures are not always compatible with the materials at hand, PECVD is a valuable alternative since the use of plasma technology and the corresponding highly reactive species enable the utilization of lower temperatures of around 150°C to 350°C, whereas LPCVD usually necessitates temperatures of around 550°C to 900°C. PECVD is therefore the commonly used approach to create passivation layers during microfabrication processes, such as either silicon oxide by reaction of silane with oxygen, [6]

$$SiH_4(g) + 2\ O_2(g) \rightarrow SiO_2(s) + 2\ H_2O(g) \tag{7}$$

silicon nitride by reaction of silane with ammonium,

$$3\ SiH_4(g) + 4\ NH_3(g) \rightarrow Si_3N_4(s) + 12\ H_2(g) \tag{8}$$

or alternating layers thereof. During the deposition, the required gases are passed over the heated substrates while a radio frequency (RF) generator induces the formation of a plasma (Fig. 7a). An important aspect of the deposited layers is their conformity and step-coverage (Fig. 7b). Since PECVD is mainly used to prepare passivation layers on top of patterned surfaces, it is of high importance to achieve good step coverage and a conformal layer which does not exhibit weak points that will result in an early breakdown of the passivation capabilities.

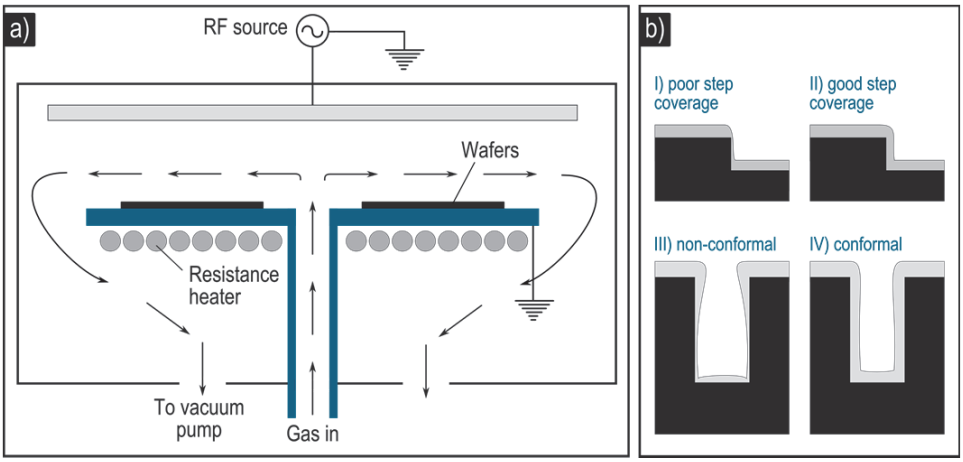


Fig. 7: *a) Schematic representation of a PECVD machine. The substrate is heated via resistance heaters while an RF source generates a plasma from the required gaseous precursors, b) layer characteristics; passivating layers need to exhibit good step coverage and conformality in order to prevent weak points (schemes according to [2]).*

Another important aspect is the presence of defects as they also present weak points in the passivation. One approach to increase the stability of the passivation is the preparation of thicker layers, however, defects often propagate through the entire deposit due to different growth characteristics in vicinity of the defects. To circumvent this problem, stacks of different materials are often used since the lattice mismatch impede the propagation of defects

from one layer to the next. In our group, multiple stacks of silicon oxide and silicon nitride are employed to solve this issue.

Atomic layer deposition (ALD) is a special case of CVD. It is a self-limiting process during which different precursors are introduced into the chamber sequentially, separated by purging steps so that only the amount of precursor A absorbed on the surface is available for reaction with precursor B, resulting in a layer-by-layer deposition of the substance. Fig. 8 depicts the processes during ALD based on the example of the formation of aluminum oxide. Trimethyl aluminum (TMA, $Al(CH_3)_3$) is introduced into the chamber and reacts with surficial hydroxyl functionalities (a) under elimination of methane (b,c). After complete surface coverage (d), unreacted TMA is removed and water is introduced into the chamber (e). As before, the chamber is purged after completion of this step. The surface functionality is now equivalent to the situation before the first cycle so that additional cycles of the process can be performed.

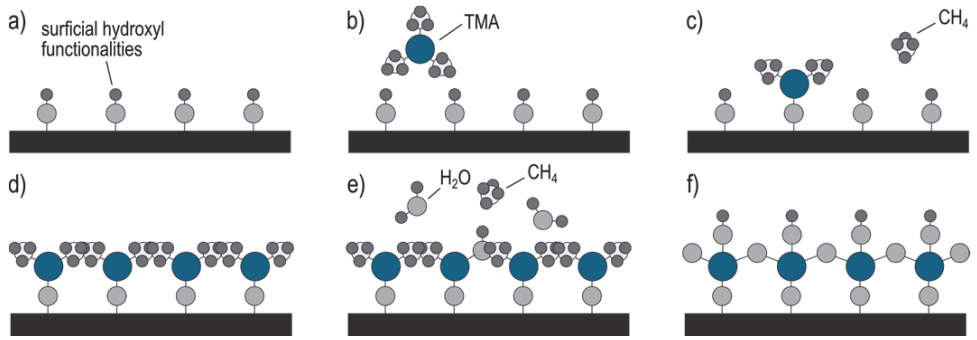


Fig. 8: *Schematic representation of the processes during ALD based on the formation of aluminum oxide. Surficial hydroxyl groups (a) react with trimethyl aluminum (TMA, $Al(CH_3)_3$) (b) under elimination of methane (c). After complete surface coverage (d), unreacted TMA is removed and water is introduced into the chamber (e). Substitution of the methyl groups under methane elimination (e) yields an aluminum oxide layer (f). After full conversion, the chamber is purged and additional cycles of the process can be performed.*

While this process certainly does not proceed in the ideal manner depicted in the schematic, layers fabricated via ALD generally exhibit fewer defects than CVD layers as well as excellent edge coverage and conformality.

Spincoating has recently become an important approach for the deposition of polymeric isolators. They offer the advantage of being processed via standard lithography techniques with exposure and development allowing for the removal of the passivation atop contact pads and electrode openings, making challenging and time consuming etching methods obsolete during this fabrication step (refer to Section 2.1 for further information on photoresist processing). However, this approach necessitates the utilization of thick layer of around 2-3µm which could have a significant influence on the contact geometry. Furthermore, many questions remain unanswered as to the long-term stability of such polymeric passivation layers in different environments and their possible hazards for biomedical applications.

3 Interfacing Electronic Devices with Neuronal Cells

Silicon-based microstructures are gaining importance in fundamental neuroscience and biomedical research. Precise and long-lasting neuro-electronic hybrid systems are at the center of research and development in this field.

For extracellular signal recordings from electrically active cells in culture, two main concepts have been developed in the past: microelectrode arrays (MEAs) (see Fig. 9a) with metalized contacts on silicon or glass substrates have been used to monitor cardiac impulse propagation from dissociated embryonic myocytes [8–10], dissociated invertebrate neurons [11,12], mammalian neurons [13], and spinal cord neurons [14]. Alternatively, arrays of field-effect transistors (FETs) (see Fig. 9b) are used for extracellular recordings having either non-metalized transistor gates with cells growing directly on the gate oxide [15–17] or metalized gates. The latter are in direct contact with the electrolyte [18] or they are electrically insulated, so-called floating gates [19–21]. With these non-invasive methods, the electrical activity of single cells and networks of neurons can be observed over an extended period of time. Meanwhile, both concepts are growing together by designing MEAs inside a CMOS process with on-chip amplification and filtering [22,23].

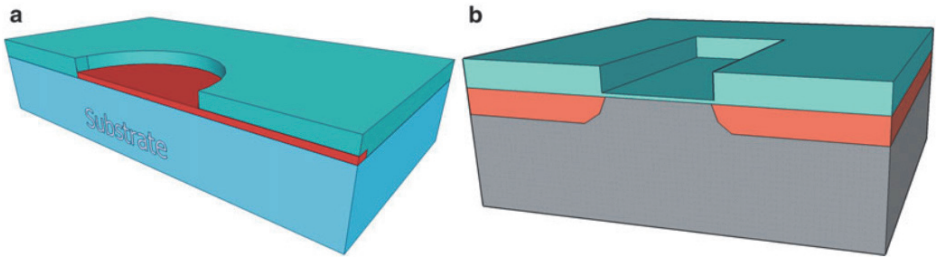


Fig. 9: *(a) Substrate-embedded microelectrode: the metal electrode (red) is exposed to the electrolyte while the feedlines are covered with an isolation layer (green-blue) (b) Open-gate field-effect transistor for the recording of extracellular signal*

3.1 Working principles

The interaction of a neuronal cell with an electronic device is schematically depicted in Fig. 10. Sufficient electrical coupling between the cell and the (gate) electrode for extracellular signal recording is achieved only when a cell or a part of a cell is located directly on top of the (gate) electrode. Electrical signals recorded by these devices show lower signals and a higher noise level (owing to a weaker coupling to the (gate) electrode) compared to intracellular electrodes or patch pipettes (see Fig. 10, right).

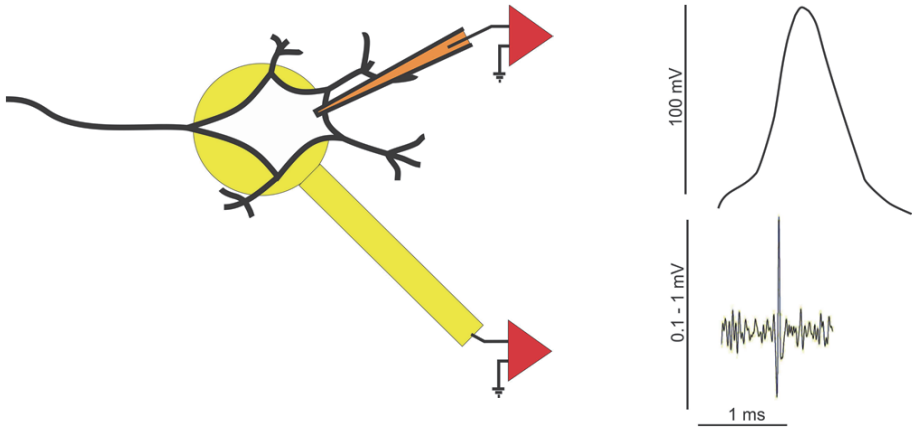


Fig. 10: *Left: Schematic of a neuron on an electronic device: intracellular (upper orange electrode) and extracellular (lower yellow electrode) signals can be recorded. Right: Action potential of a neuron (approx. 100 mV) recorded by an intracellular electrode (upper trace) and an extracellular electrode (lower trace)*

For a quantitative understanding of the extracellular signals recorded by electronic devices, it is necessary to explain the experimental situation in detail. A schematic picture of a typical experimental situation is depicted in Fig. 11. Here, the neuroelectronic hybrid is formed by the neuron, the cleft between neuron and the sensor surface, and the electronic device. Outside the neuron and inside the cleft, there is extracellular electrolyte solution. By electrical excitation, the ion channels in the cell's membrane open and ions can flow from across the cell membrane.

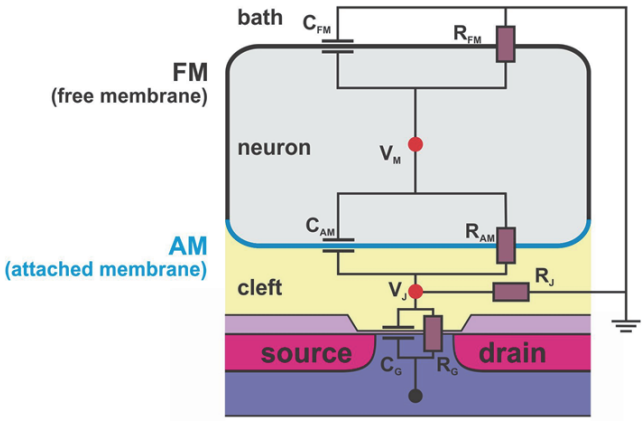


Fig. 11: *Schematic of the neuroelectronic hybrid. The cell membrane is divided into free (FM) and attached membrane (AM) with the respective values of membrane area (A_{FM}, A_{JM}) and membrane capacitance (C_{FM}, C_{JM}) and resistance (R_{FM}, R_{JM}). C_G and R_G are the capacitance and the resistance of the (gate) electrode, respectively. The seal resistor R_J represents the electrical properties of the cleft between the membrane and the sensor surface. In case of patchclamp experiments, the intracellular voltage V_M can be determined.*

While in the upper part of the cell (free membrane) these ions just enter the surrounding electrolyte bath directly, it is different at the attached membrane. Here, the ions have to pass the cleft before entering/leaving the bath. The cleft acts as a resistance typically called seal resistance R_J [16,24]. The magnitude of R_J is typically in the order of several 100 kΩ up to MΩ corresponding to a typical cleft thickness of 40 to 150 nm [25–27]. The voltage V_J, which determines the voltage at the (gate) electrode, is mainly determined by the seal resistance R_J and the current that flows across it.

3.2 Planar Microelectrode Arrays

Planar MEAs were first reported by Thomas et al. in 1972 [8]. The aim of their research was to establish an electrophysiology technique that enables the parallel, non-invasive recording from electrically active tissues and cell cultures over periods of days to weeks, a goal which was not achievable with alternative methods at the time such as sharp microelectrodes. Their design consisted of 30 electrodes arranged in two lines 50µm apart, a distance of 100µm between the electrodes within a line and 50µm² area of each electrode (Fig. 12).

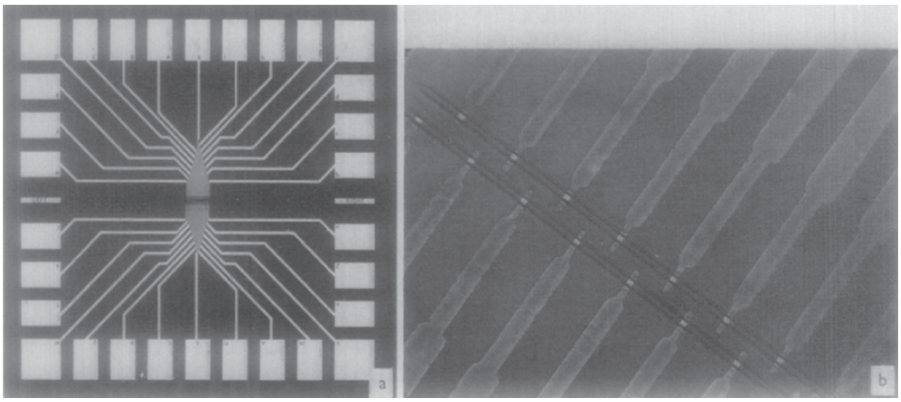


Fig. 12: *First MEA as developed by Thomas et al. in 1972. Their design consisted of 30 electrodes arranged in two lines 50µm apart, a distance of 100µm between the electrodes within a line and 50µm² area of each electrode (image taken from [8]).*

They deposited platinum black on each electrode and succeeded in recording action potentials of up to 2.5mV in amplitude from confluent, contracting layers of embryonic chick cardiomyocytes. Since then, various designs and materials have been employed for the fabrication of novel devices for improved signal quality and stability. While our group investigates a variety of different design, our most common layout exhibits 64 electrodes arranged in an 8x8 grid (Fig. 13).

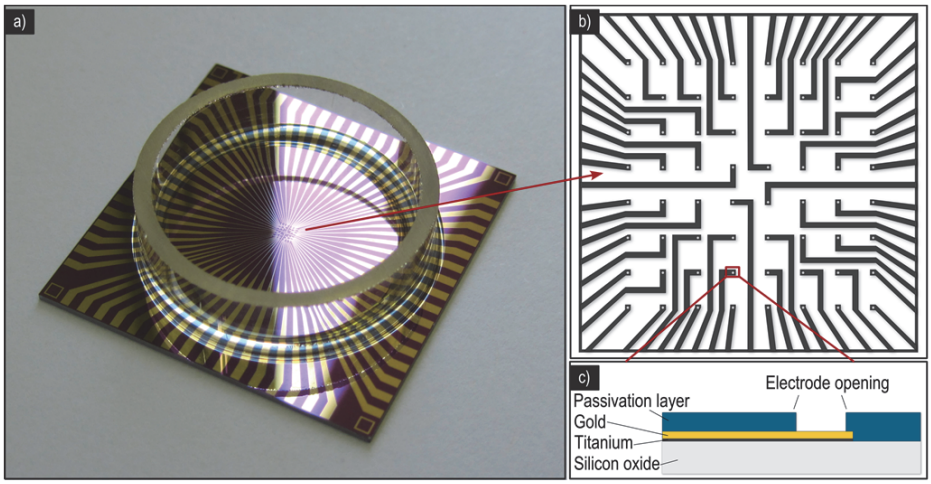


Fig. 13: *64 electrode MEA layout. a) 64 electrode chip with attached glass ring as fluid reservoir. b) Schematic of the electrode array showing the arrangement of the 64 electrodes. c) Schematic cross-section of the electrode. Titanium is used as adhesion layer between the silicon oxide substrate and the gold electrode. An opening in the silicon oxide/silicon nitride passivation layer allows contact between the environment and the electrode.*

For the fabrication of our devices we use standard cleanroom technology as described in Section 2. In Fig. 14, the process is described for an exemplary positive photoresist. Starting from a plain silicon wafer (Fig. 14a), an oxidation is conducted, forming an insulating layer on top of the semiconductor (Fig. 14b). Afterwards, photoresist is spincoated onto the substrate and then exposed through a chromium mask (Fig. 14c). This determines the pattern transferred onto the photoresist, which, in this particular case, are the 64 electrodes and corresponding feedlines of the MEA (Fig. 14d). During exposure, the polymeric photoresist decomposes in illuminated areas (Fig. 14e) and is subsequently removed during development (Fig. 14f). Metal is evaporated onto the pattern, covering both the remaining photoresist and the bare silicon oxide (Fig. 14g). The metal or stack of metals and the corresponding layer thicknesses are dependent on the final application and can be tailored accordingly by changing the parameters of the PVD process. After the metal deposition, the remaining photoresist is removed in acetone, which, consequently, also removes metal layers deposited on top of the photoresist (lift-off), leaving a structured metal surface (Fig. 14h).

The surface is then passivated either via PECVD, ALD or spincoating of a polymeric layer to form an insulation between feedlines and the environment (Fig. 14i), with the most commonly used passivation being alternating stacks of silicon oxide and silicon nitride as prepared by PECVD. Subsequently, a second lithography is performed (Fig. 14j) to open apertures in the center of the chip as well as bond-pads in the periphery that enable the chip to be contacted to the external electronics. After exposure (Fig. 14k) and development (Fig. 14l), the photoresist serves as mask for a RIE process which enables the local removal of the passivation layer and uncoverage of the metal electrode area (Fig. 14m). Afterwards, the dissolution of the photoresist yields the finished planar MEA device (Fig. 14n).

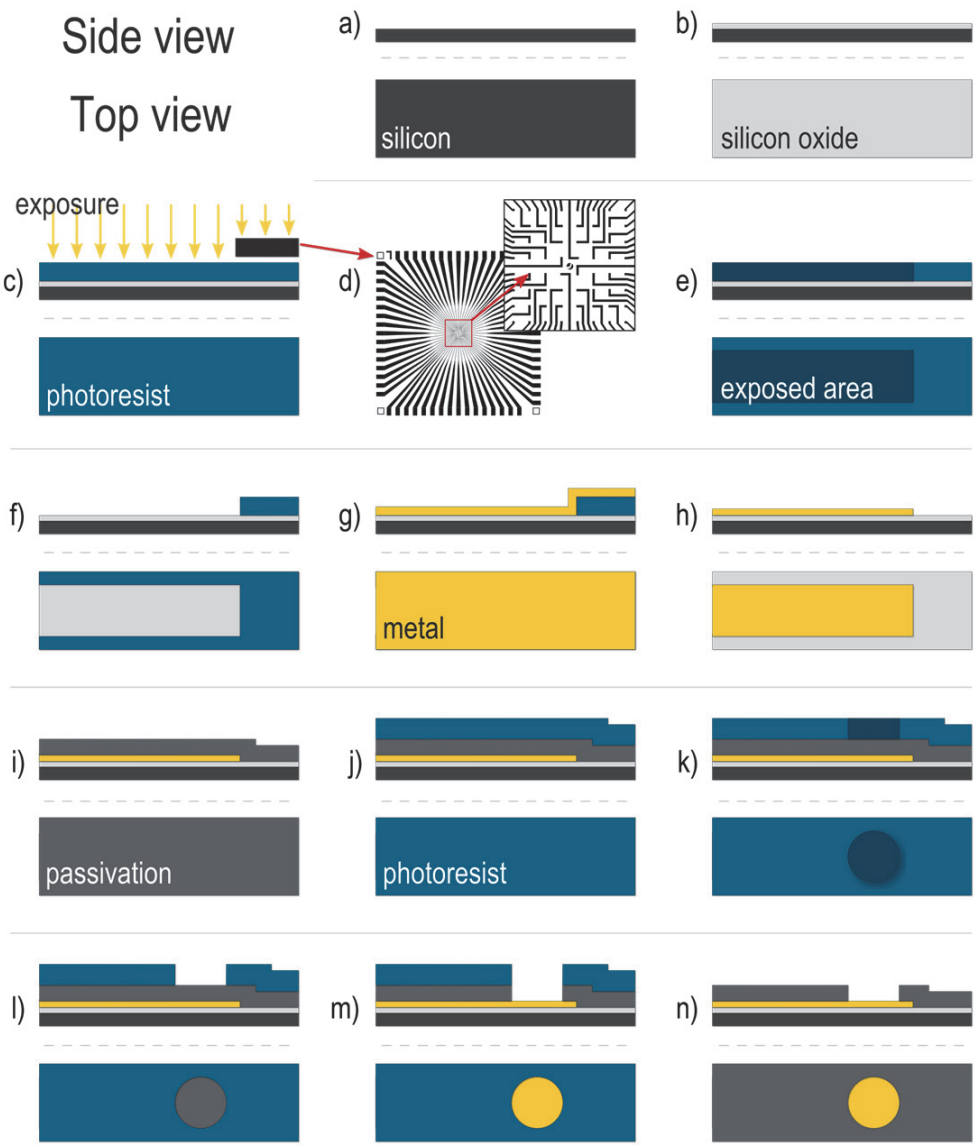


Fig. 14: *Fabrication of MEA by example of a single electrode. Silicon (a) is oxidized to form an insulating oxide layer (b). Photolithography (c-f), PVD (g) and lift-off (h) enable the formation of electrodes, which are then passivated via PECVD (i). A second lithography step (j,k) opens apertures (l) which enable a localized RIE etching of the passivation, presenting an active area of the metal electrode (m). Photoresist removal yields the final device (n). Note that the mask in (d) depicts the pattern for the entire 64 electrodes.*

While many improvements of planar MEAs have been developed over the years, ranging from increasing the number and density of electrodes, introducing high surface areas via

platinum black or chemical modification to increase cell adhesion and thus decrease the gap between cell and electrode, planar MEAs still suffer from limited signal quality. Furthermore, it remains impossible to record subthreshold potentials using this method [28]. Due to this reason, a variety of more complex designs based on the idea of extracellular, non-invasive MEAs have been developed. The following sections will provide information on such possible improvements.

3.3 3D Electrodes

Due to the fundamental limitations of planar MEAs, various modifications have been introduced in the literature to facilitate a stable interface for efficient cell-chip coupling that enables robust and long-term electrophysiological recording and stimulation [28]. One possible approach is the introduction of three-dimensional features as suggested by Spira et al. in 2007 [29]. They designed gold mushroom-shaped microspines based on the idea of exploiting the fundamental cellular process of phagocytosis (see Fig. 15), which enables the internalization of larger objects, and could therefore enable a closer contact between cell and electrode and hence an increase in seal resistance and thus signal quality.

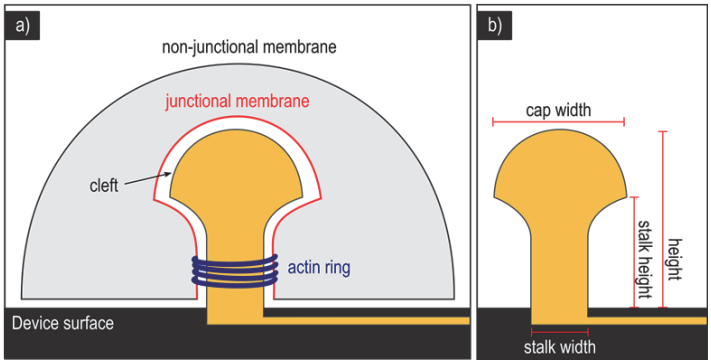


Fig. 15: *a) Schematic of the proposed mechanism for the improvement of cell-electrode contact. Phagocytosis could lead to a reduction in cleft size between cell and electrode. Actin ring formation around the mushroom stalk further supports this mechanism. b) Geometrical parameters of mushroom-shaped 3D electrodes (adapted from [30]).*

Spira et al. have continued extensive research on their idea of 3D mushroom-shaped electrodes. After their first experiments in 2007, which showed that both aplysia neurons and cells from a human cardiomyocyte cell line readily engulf gold mushroom-shaped structures of around 850nm stalk width, 1µm stalk height, 1.8µm cap width and 1.6µm total height [29], they proved the formation of an actin ring around the stalks for cultures of aplysia neurons [31]. Subsequently, they performed first electrical recordings from aplysia neurons on microspines and compared them to recordings on planar electrodes [32]. They found that while single spines yielded recordings similar to that of planar electrodes, an assembly of four gold spines allowed for a higher than 4 fold increase in signal as compared to planar electrodes. Since then, many different ideas for 3D electrode structures have been suggested in the literature.

In our group, we investigate various different 3D approaches for the improvement of MEA technologies. One method is the employment of nanoporous alumina templates which are

prepared on the devices via self-assembly, followed by electrodeposition of gold inside the template for the formation of gold nanopillar arrays (Fig. 16).

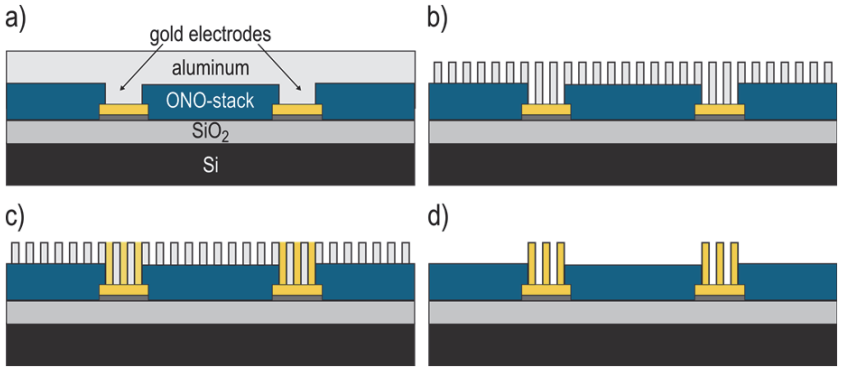


Fig. 16: *Preparation of porous alumina templated gold nanopillar MEAs. The aluminum layer (a) is anodized, preparing a regular structure of nanoholes (b). Since only areas above the electrodes are in contact with the electroplating solution, gold is selectively deposited in the electrode area. Removal of the aluminum oxide yields free-standing gold nanopillars (adapted from [33]).*

The resulting dense arrays of gold nanopillars (Fig. 17) could be shown to result in a significant reduction in electrode impedance and were employed by Brüggemann et al. [33] to perform action potential recordings from cardiomyocyte-like HL-1 cell line.

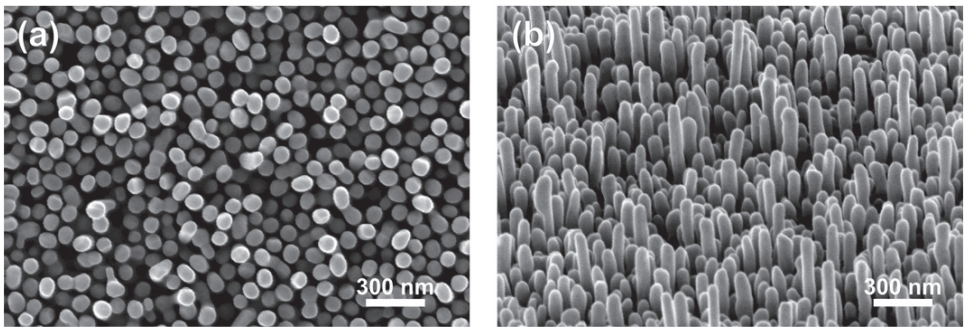


Fig. 17: *Top (a) and side view (b) of gold nanopillar arrays with pillars of 300 to 400nm in height and diameters of approximately 60nm as fabricated using a self-assembled template of nanoporous aluminum oxide (image taken from [34]).*

They were able to demonstrate good cell survival on the structures and recorded action potentials of up to 1.5mV peak-to-peak for electrodes of 20µm in diameter, with the introduction of gold nanopillars resulting in up to a two-fold increase in signal amplitude. Focused ion beam (FIB) cross-sections of HL-1 cells grown on nanopillars proved the formation of a tight contact between cell and electrode (Fig. 18).

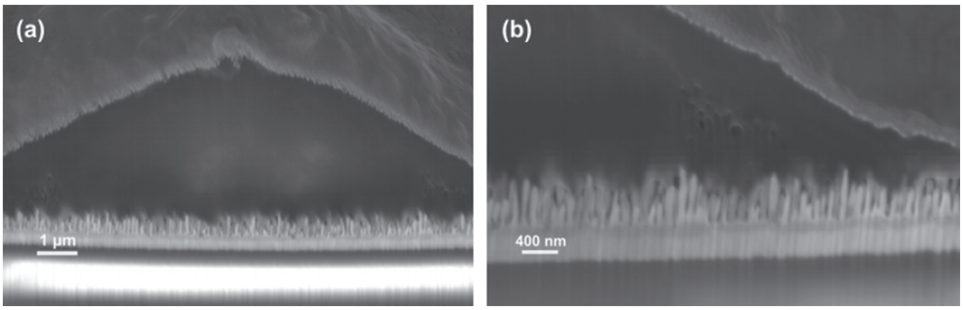


Fig. 18: *Focussed ion beam (FIB) cross-section of an HL-1 cell grown on gold nanopillar arrays at different magnification. The close contact between cell and electrode facilitates good electrical coupling (image taken from [33]).*

In addition to the templated nanopillar arrays, we also investigate gold mushroom-shaped 3D electrodes, based on the findings of Spira et al. These structures are prepared via electron-beam (e-beam) lithography [35] as depicted in Fig. 19. An e-beam sensitive polymer such as poly(methyl methacrylate) (PMMA) is spincoated onto a gold covered carrier (Fig. 19b) and hole patterns are introduced into the layer via e-beam exposure (Fig. 19c). Gold is then electrodeposited into the holes, yielding either pillars (Fig. 19d_1) or mushroom-shaped electrodes (Fig. 19e_2) depending on whether the galvanization is allowed to proceed past the boundaries of the template. Subsequent removal of PMMA yields the free-standing structures (Fig. 19e_1/f_2). A heating step introduced before the galvanization enables rounded corners and thus a more biomimetic shape of the mushroom-shaped electrodes (Fig. 19d_2).

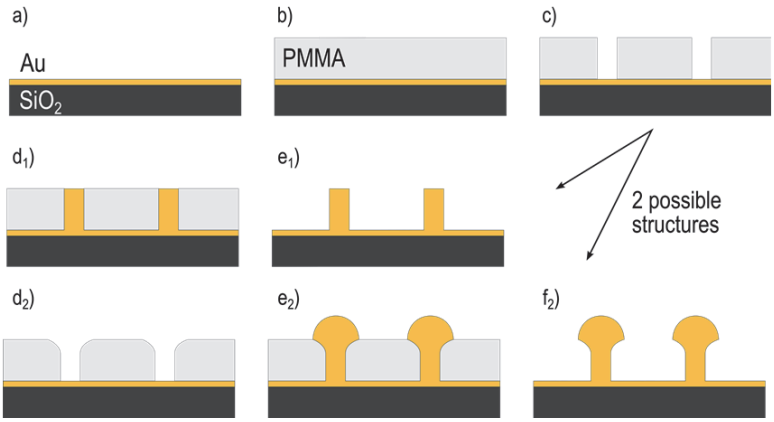


Fig. 19: *Schematic representation of 3D electrode fabrication. PMMA is spincoated onto a gold layer (b) and pattern via e-beam exposure (c). Gold is then electroplated into the template, yielding either pillars (d_1) or mushroom-shaped electrodes (e_2) depending on whether the galvanization is allowed to proceed past the boundaries of the template. Heating of the PMMA before electroplating results in rounded corners and thus a more biomimetic shape of the mushroom-shaped electrodes (d_2). PMMA removal yields the free-standing electrodes (e_1/f_2).*

Fig. 20 shows scanning electron microscopy (SEM) images of the resulting mushroom-shaped structures.

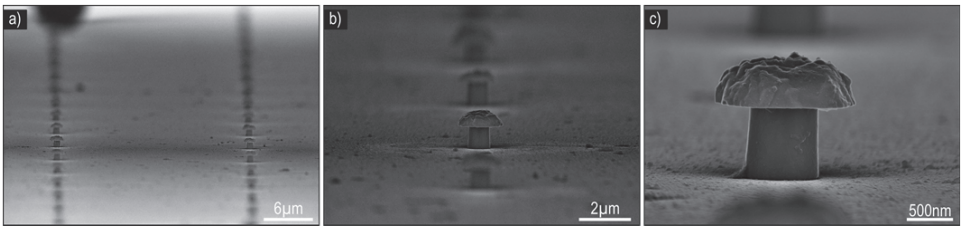


Fig. 20: *SEM images of 3D mushroom-shaped electrodes.*

We have performed extensive SEM studies and FIB cross-sections of cells interacting with 3D electrodes. It could be shown that high aspect ratios between stalk height and width result in better cell-electrode contact [36] and that the "cap" on mushroom-shaped structures results in a better engulfment by the cell as compared to plain pillars. Fig. 21 shows SEM images of HL-1 cells both on pillars and mushroom-shaped 3D structures.

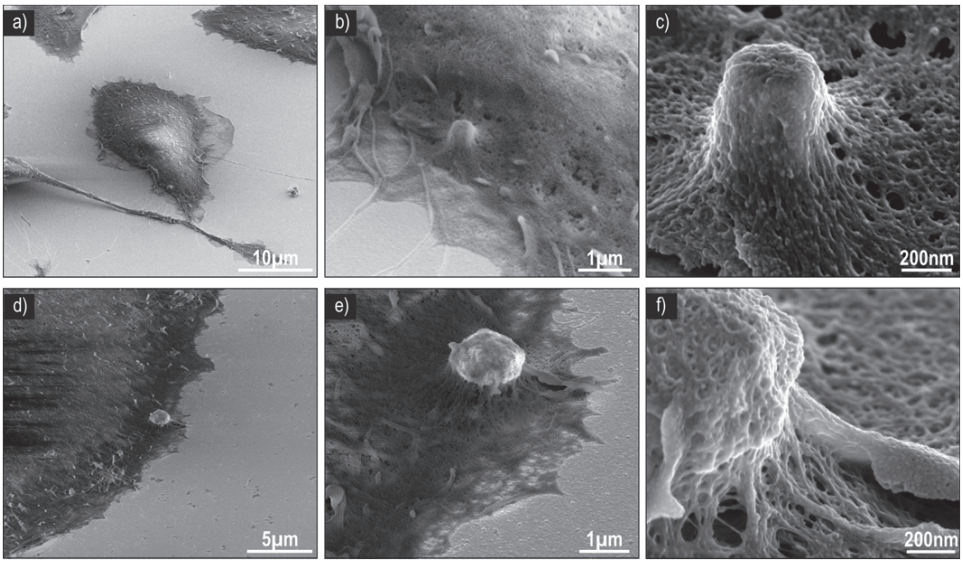


Fig. 21: *SEM images of HL-1 cells on pillars (a-c) and mushroom-shaped 3D structures (d-e). The cells grow on top of the 3D structures in both cases but while plain pillars lead to a tent-like formation, mushroom-shaped structures induce cellular engulfment of the 3D electrode (image taken from [36]).*

FIB cross-sections [37] confirmed a tight contact between cell and mushroom-shaped 3D structures (Fig. 22).

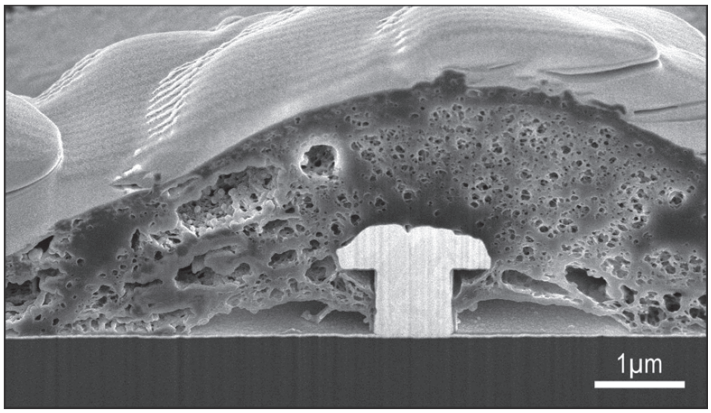


Fig. 22: *Focussed ion beam cross-section of an HL-1 cell engulfing a mushroom-shaped 3D electrode. The observed tight contact between cell and electrode facilitates good electrical coupling (image courtesy of F. Santoro).*

The transfer of these structures onto planar MEAs enabled the recording of action potentials from HL-1 cells. Assemblies of nine mushroom-shaped electrodes were employed for the recording, yielding signals of an amplitude of up to 180µV [38], which is comparable to the results obtained by Xie et al. [39].

Yet another approach which is investigated in our group is the usage of nanocavities [40,41]. They can be considered to be an "inverted" version of 3D electrodes, enabling not an engulfment by the cell but rather a protrusion of the cell into the sensor. These structures are prepared using a stack of either platinum or gold and chromium as electrode material (Fig. 23a). Small apertures of up to 24µm are etched through the passivation layer via reactive ion etching (Fig. 23d). Since chromium can easily be dissolved using commercial etching solutions which leave both platinum and gold layers intact, a selective underetching of the passivation by removal of chromium enables the formation of a nanocavity (Fig. 23f).

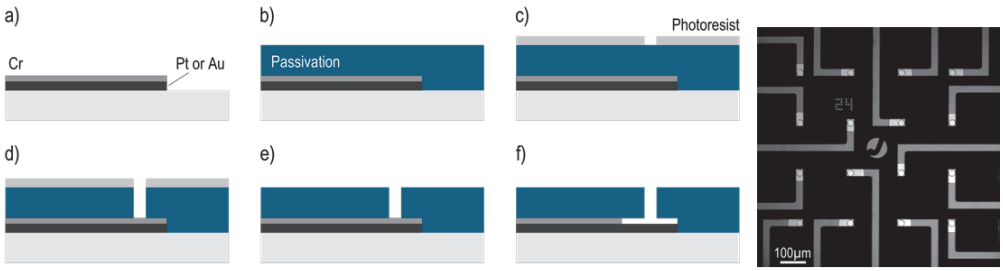


Fig. 23: *Left: Schematic of nanocavity fabrication. The electrode material is covered by a sacrificial layer of chromium (a). The device is then passivated (b) and small apertures of up to 24µm in diameter are introduced via lithography coupled with reactive ion etching (c-e). Removal of the chromium via an etching step results in the formation of a nanocavity (f) (scheme according to [41]). Right: Nanocavity chip after etching. The bright areas surrounding the circular apertures mark the underetched area.*

While the vertical dimension of the cavitity can be tailored via the thickness of the chromium layer, the horizontal dimension depends on the parameters of the etching process. Due to the increase of the active surface area, the impedance of the device decreases dramatically, which is an important aspect for low-noise electrophysiological measurements [42]. Recordings from HL-1 cells using these devices exhibit stable action potentials of an amplitude of around 1mV with excellent signal-to-noise characteristics [41]. Furthermore, this approach also enables the stimulation of electrical activity [42].

In future studies it would be interesting to combine 3D structures with geometries that enable cellular protrusion into the active area to merge the advantages of 3D electrodes and nanocavities. This could be achieved via the preparation of hollow pillars/nanohollows (Fig. 24).

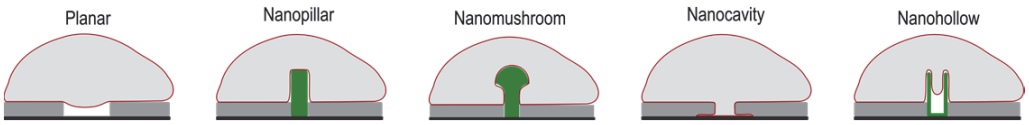


Fig. 24: *Evolution of electrode designs from planar MEAs to complex electrode structures for improved cell-chip coupling.*

Recent developments utilizing this approach have shown promising results [43], presenting a good basis for further investigations.

3.4 Fieldeffect Transistor Arrays

P. Bergveld introduced in 1970 a field-effect transistor (FET) as transducers for (electro-/bio-) chemical sensing [15]. Here, the gate metal electrode of the metal-oxide-semiconductor FET (MOSFET) was replaced by an electrolyte solution. Changes in the surface charge of the FET due to ionic or molecular interactions or due to extracellular potentials lead to an additional gating of the conductive channel of the FET and is converted into an electrical signal.

For a theoretical description of the operation of ion-sensitive field-effect transistors (ISFET), the voltage drops across the reference electrode, the electrolyte solution, the surface potential at the interface between electrolyte and gate oxide, the gate oxide and part of the silicon needs to be considered. They depend on the specific doping level of the silicon, the gate oxide charge, the composition of the electrolyte solution and the reference electrode type.

Silicon nanowire (SiNW) field-effect transistors (Figure 17a) seem to be a promising candidate for the down-scaling of the ISFET concept to the nanoscale which was firstly demonstrated by Lieber and coworkers [44]. Beside the chemical or biomolecular detection, the SiNW devices can also be used to study the activity of electrogenic cells [45–47].

Methods for SiNW fabrication can be divided into two categories: a bottom-up approach that is usually based on vapor-liquid-solid method (VLS) with metal precursors, mostly Au, and a top-down approach that is based on lithography techniques. In the latter case usually silicon on insulator (SOI) substrates are used.

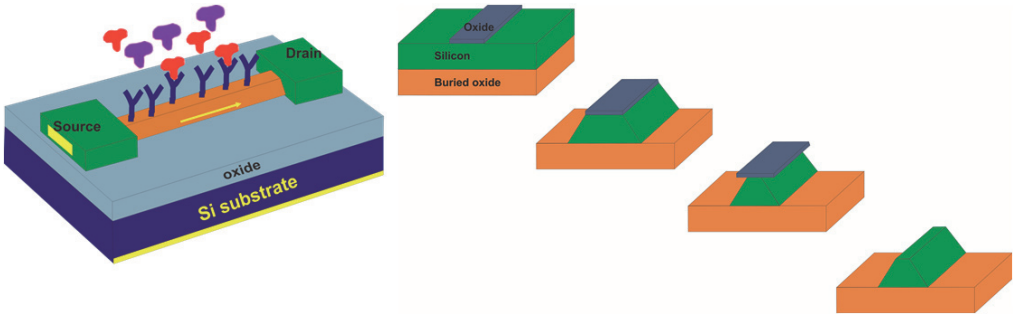


Fig. 25: *Left: schematics of the sensing concept using SiNW-FET. Right: schematics of the 'Top-down' fabrication process of SiNWs by etching of Si on a silicon on insulator (SOI) wafer. Trapezoidal-shaped nanowire cross section is caused by the anisotropic wet etching of silicon (TMAH etching process). The height of the wires is given by the top Si layer thickness of the SOI-wafer.*

Although the bottom-up method is a very straight-forward way to produce nanowires, the top-down approach has several advantages: the nanowires are uniform in size and well aligned in predetermined orientation and position on the substrate. SiNWs can be produced by top-down approach with widths down to 50 nm and lengths ranging from 3 µm up to 1 mm. Dimensions similar to the bottom-up approach can be achieved by size reduction in a process called self-limiting oxidation after structuring or by wet anisotropic etching of Si [48,49] (see Figure 18b).

The structure of the SiNW can be analysed by scanning transmission electron microscopy (STEM). The analysis indicates non-ideal trapezoid shape and different gate oxide thickness (7 nm on the (100) plane and 13 nm on the (111) plane). The underetching of the nanowire is originating from several wet etching steps during fabrication. These devices were used to record the extracellular potential of the spontaneous activity of cardiac muscle HL-1 cells (Fig. 26). Their signals were measured by direct dc sampling of the drain current. An improved signal-to-noise ratio compared to planar field-effect devices was observed. Furthermore the signal shape was evaluated and could be associated to different membrane currents.

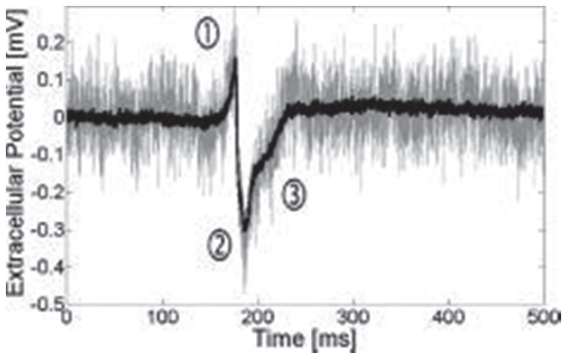


Fig. 26: *Representative extracellular recording of an HL-1 action potential with a planar FET (black: average of 57 action potentials, gray: single trace).*

3.5 High-Density CMOS-type MEA arrays

High spatial resolution recording is desirable for the studies of cell activity in neural circuits. Spatial resolution and number of electrode can be increased by an ultra-high density MEA with thousands of channels using high-density VLSI circuit design technology. In addition, this will allow the integration of dense microelectrodes, amplifiers, filters, stimulation buffer, multiplexer, digital logic circuits, and analog-to-digital converter circuitry on the same silicon substrate. The CMOS type MEA can be fabricated by combining VLSI circuit fabrication on a chip followed by the microelectrode array process. During this post-processing, electrodes are fabricated and passivation layers are added to protect the active circuitry from electrolytes.

Several research groups have developed CMOS-type MEAs with high density electrode arrays. Fromherz and coworkers used 0.5-μm CMOS process to make a 128 × 128 electrode sensor chip (sensing area: 1 mm × 1 mm) [50]. The individual pixel was realized by an oxide-semiconductor field-effect transistor with a footprint of 7.8 μm × 7.8 μm (Fig. 27a). The final layer system consisted of Ti and Pt and sputtered 40 nm stacks of TiO_2 and ZrO_2 (Fig. 27b). With this design they achieved a 2 kps full frame rate for recording from the entire array. However, the power consumption was very high and a temperature regulation of the chip for biological experiments was necessary. The group reported improved noise level (40 ~ 80 μV) and faster frame rate (6 kHz) in an next chip generation which was applied to study the signal propagations in retinal ganglion networks [51].

Berdondini and coworkers applied a 0.35 μm CMOS process to design a 64 × 64 active pixel electrode MEA (active area: 2.67 mm × 2.67mm) [23]. As a sensing element they used gold electrodes (21 μm × 21 μm, spaced by 21 μm). To handle the large amount of data, a frame grabber was used. Full frame rate was 7.8 kHz for 4096 channels and input referred noise level was 11 μVrms. The power consumption was 132 mW, which did not raise the temperature of the cell culture media significantly. The chip was used in experiments with rat hippocampal and cortical neurons, acute brain slices [52], and mouse retinal tissues.
Hierleman and coworkers used 0.6 μm CMOS process for a high density MEA electrode system (sensor area: 2.0mm × 1.75mm) with the capability of simultaneous recording and stimulation of 126 channels [22,53] (Fig. 27d). Each electrode had the size of 7 μm in diameter and electrodes were spaced by 18 μm. To lower the electrode impedance, platinum black was deposited on the Pt electrodes (Fig. 27e).

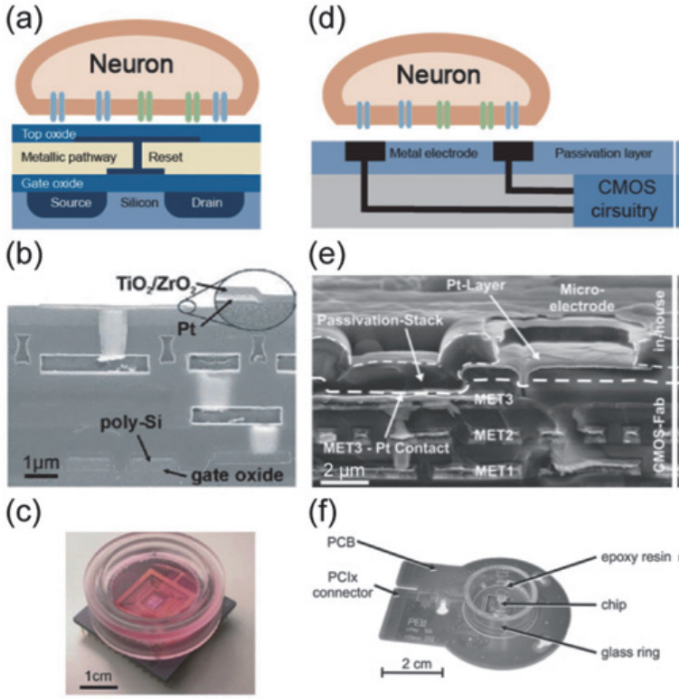


Fig. 27: *Examples of CMOS-type MEAs. Multi-transistor array developed by Fromherz and coworkers used the oxide layer as sensing element (a). The device structure (b) and a packaged chip (c) are shown. Microelectrode array integrated with CMOS circuitry developed by Hierlemann and coworkers used metal electrode as sensing element (d). The device structure (e) and a packaged chip (f) are shown (figure from Y. Nam, Biomedical Engineering Letters 4 (2014), 129-141).*

The input referred noise level was 2.4 µVrms (1 Hz – 100 kHz) and the power consumption was 135 mW. This chip was used to investigate the extracellular electrical field in Parasagittal cerebellar slices. More recently, Bakkum and coworkers combined the on-chip electrical stimulation capability and high-resolution recordings to map the propagation of action potentials in a single cortical neuron [54]. This group recently reported an upgraded version of the chip capable of recording 1024 channels from 26400 electrodes.

4 Patterning neuronal networks on electrode arrays

Beside the improvement of the cell-electrode coupling by modifying the electrode surface, the capturing of a cell on the electrode is a prerequisite for the recording of a neuronal signal with an electrode array. To achieve a good coupling between the cell and the electrode at least a close proximity of the cell to the electrode surface is required. For homogeneously growing cell cultures this coupling mainly depends on the cell density on the substrate. A low cell density on the electrode array results in a low number of electrodes covered with cells and thus poor signal recordings. Culturing a high number of cells on the chip, however, suffers from the loss of the ability to assign and track single cell signals. Therefore, it is advantageous to pattern the cells directly on the electrodes to improve the cell electrode coupling. Neuronal

outgrowth responds to different factors like intrinsic cues such as centrosome position [55], or extrinsic cues made up of extracellular signaling proteins [56,57] or topological structures. Inspired by the variety of ways to influence cell growth, numerous approaches to pattern cells on an electrode array arose during the last years. The employed patterning methods can be separated in two mayor categories: Controlling individual neurons, and neuronal populations.

4.1 Controlling adhesion and outgrowth of individual neurons

As a first step of improvement, cell adhesion on the chip surface must be facilitated, as most of the employed top layer materials of an electrode array do not promote cell adhesion. The most common technique of supporting cell growth on chips are surface chemical modifications such as coating with extracellular matrix proteins like poly-L-lysine, poly-D-lysine, laminin, or hydrogels [58–61]. These modifications can be performed homogeneously on the whole chip or spatially controlled. As neurons tend to migrate after first adhesion on the surface [62,63], a specific attractive force of the electrodes to the soma needs to be achieved. This can be reached by introducing cell attractive in-mold structures [64], spatially restricting the growth of the soma simply by the available surface for adhesion. The same attempt restricts neurite outgrowth by introducing small passages guiding the neurites between the soma.

Less constricting methods arose from approaches using microcontact printing to structure cell repellant surfaces with cell adhesive proteins. Differently shaped adhesion points were investigated with regards on axonal outgrowth [60,65] and the results indicate that the nucleus position affects the orientation of the sprouting axon. Nevertheless, controlled soma position allows the generation of networks of neurons on an electrode array with different geometries [58,59,66]. The polarity of neuronal networks can be influenced on the level of individual neurons by introducing connections between soma adhesion points with a nonlinear design [60,61,67]. This approach offers the opportunity to create networks with predefined signal pathways, with full accessibility of single cells for external modification.

4.2 Controlling connectivity of neuronal populations

Culturing populations of neurons on electrode arrays suffers from the lack of controllability of the contact between cells and the electrodes, as the cells are generally homogeneously distributed within the population. To increase the probability of cells located on the electrode, nanopillars can be introduced on the electrode surface [68], which reduces the cell mobility from 57.8µm to 3.9µm over a period of 5 days. However, with this improvement there is still only little control over the connectivity of neuronal populations. Approaches comparable to those described in Section 4.1 for individual neurons were made. In-mold patterning of neuronal populations was used to control their connectivity [69,70]. Using this method, a directionality of signal propagation of over 90% among the populations of neurons was achieved.

Also, logic devices have been generated by this group allowing numerical operations on an electrode array [70]. Promising other approaches came up by Peyrin and co-workers [71], who introduced a microfluidic system connecting different populations of neurons. By geometrically shaping the microchannels, they are able to specifically guide axons from one populations to the other without axons growing in the opposite direction. By aligning the microfluidic system to the electrode array, signals could be recorded from the different neuronal populations and even from the sprouting axons in the microfluidic device. This

system allows for the investigation of neuroregeneration after lesion by shear forces at an additional channel perpendicular to the microfluidic channels [72]. However, while enabling good control over the network connectivity, the system suffers from the lack of accessibility of single cells to external manipulation. To overcome this downside, microcontact printing can be used to control the connectivity of adjacent neuronal populations (Fig. 28) [73].

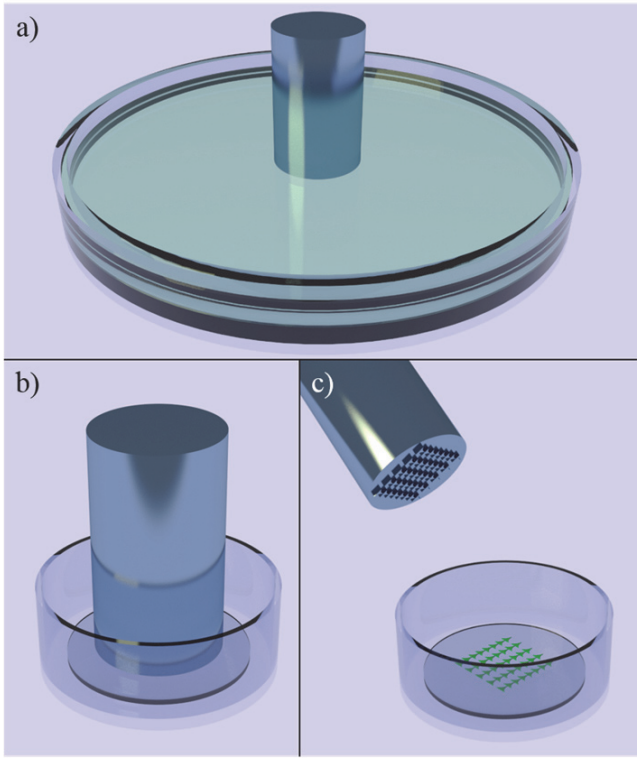


Fig. 28: *Microcontact printing process for the formation of protein patterns. a) Inking of stamp and drying in N_2-stream. b) Stamping on substrate results in the transfer of the protein. c) After stamping, the protein remains on the substrate.*

Fig. 29 shows neurons growing on printed triangular pattern with neurites sprouting randomly. As it has been shown by Albers and colleagues, connectivity heavily depends on the shape of the gateway between adjacent populations. Choosing the right geometric parameter, mainly characterized by the opening angle of the triangle, neurites can be funneled from one population towards the other, leading to controlled connectivity between adjacent populations of neurons. This can be verified by recording neuronal activity via calcium imaging.

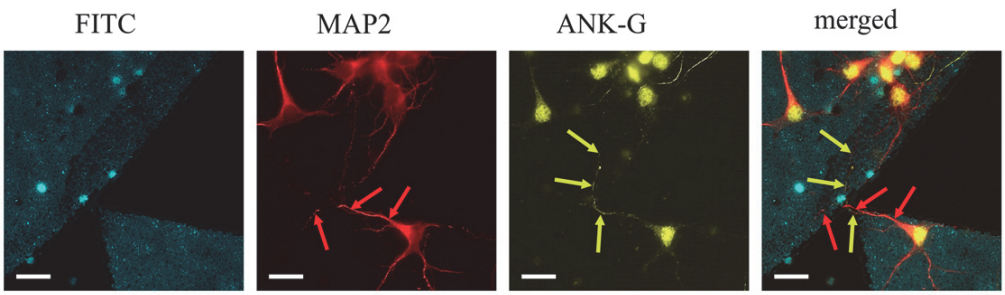


Fig. 29: *Neuronal populations growing on triangular pattern described in more detail by Albers and colleagues [73]. The microcontact printed protein is FITC labeled and can thus be visualized via fluorescence microscopy. The neurites develop well and small branches can be seen by MAP2 staining. The axons were stained with ANK-G to track their development. In the merged image, the position of the axon and the dendrites are indicated with the yellow and red arrows introduced in the ANK-G and MAP2 images. The neurites are funneled towards the apex of the triangular structure. Scale bar: 20 µm.*

A representative image of propagating action potentials indicated by the delay of their occurrence is shown in Fig. 30.

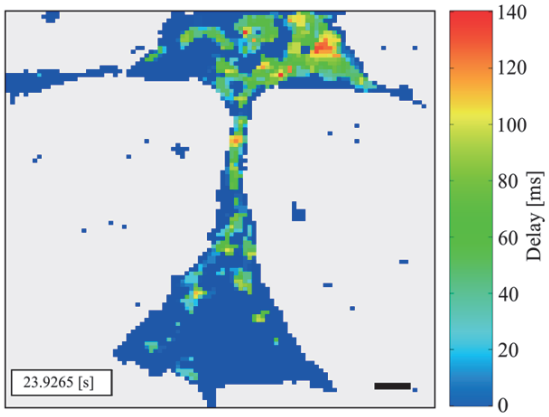


Fig. 30: *Plot of delayed occurrence of neuronal activity. Action potentials (APs) were recorded by calcium imaging. Changes of fluorescence intensity were detected and correlated accordingly. The activity arises as indicated by the colorbar, whereas the first AP showed up at the time indicated at the insert at the bottom left. Scale bar: 50 µm.*

However, there is still no specific patterning of single cells on the electrodes. The cells grow homogeneously distributed within the neuronal population and neurons adhere directly on an electrode only by chance. To overcome this limitation, the NeuroArray system was introduced [74]. Here, circular holes are molded into a PDMS layer from an SU8 stencil. The holes are arranged in a triangular shape comparable to the design and dimensions Feinerman and colleagues introduced in 2008 [70]. Neurons were cultured on the PDMS layer and the

soma positions itself in the holes. By adding small bumps to the top side of the PDMS neurites can develop freely on either sides of the PDMS by turning it upside down in a culture dish. As the soma of the neurons is localized in the holes they can easily be aligned to the electrodes and the number of cells on an electrode can be tuned by varying the hole diameter.

References

[1] McCord and Rooks, *SPIE Handbook of Microlithography, Micromachining, and Microfabrication* (Society of Photo-Optical Instrumentation Engineers, 1997).
[2] B. Bhushan, *Springer Handbook of Nanotechnology* (Springer, 2010).
[3] Z. Cui, *Nanofabrication Principles, Capabilities and Limits* (Springer, 2008).
[4] P. A. Kohl, in *Mod. Electroplat.* (John Wiley & Sons, Inc., 2010), pp. 115–130.
[5] T. I. K. Richard S. Muller, *Device Electronics for Integrated Circuits*, 3rd ed. (WILEY-VCH Verlag, 2003).
[6] J. P. Uyemura, *Introduction to VLSI Circuits and Systems* (WILEY-VCH Verlag, 2002).
[7] R. J. Martin-Palma, *Nanotechnology: A Crash Course* (SPIE Press, 2010).
[8] C. A. Thomas, P. A. Springer, G. E. Loeb, Y. Berwald-Netter, and L. M. Okun, Exp. Cell Res. **74**, 61 (1972).
[9] D. A. Israel, W. H. Barry, D. J. Edell, and R. G. Mark, Am. J. Physiol. **247**, H669 (1984).
[10] P. Connolly, P. Clark, A. S. Curtis, J. A. Dow, and C. D. Wilkinson, Biosens. Bioelectron. **5**, 223 (1990).
[11] L. J. Breckenridge, R. J. Wilson, P. Connolly, A. S. Curtis, J. A. Dow, S. E. Blackshaw, and C. D. Wilkinson, J. Neurosci. Res. **42**, 266 (1995).
[12] W. G. Regehr, J. Pine, and D. B. Rutledge, IEEE Trans. Biomed. Eng. **35**, 1023 (1988).
[13] J. Pine, J. Neurosci. Methods **2**, 19 (1980).
[14] G. W. Gross, A. N. Williams, and J. H. Lucas, J. Neurosci. Methods **5**, 13 (1982).
[15] P. Bergveld, J. Wiersma, and H. Meertens, IEEE Trans. Biomed. Eng. **23**, 136 (1976).
[16] P. Fromherz, A. Offenhäusser, T. Vetter, and J. Weis, Science **252**, 1290 (1991).
[17] A. Offenhäusser, C. Sprössler, M. Matsuzawa, and W. Knoll, Biosens. Bioelectron. **12**, 819 (1997).
[18] D. T. Jobling, J. G. Smith, and H. V Wheal, Med Biol Eng Comput **19**, 553 (1981).
[19] A. Offenhäusser, J. Rühe, and W. Knoll, J. Vac. Sci. Technol. a **13**, 2606 (1995).
[20] A. Cohen, M. E. Spira, S. Yitshaik, G. Borghs, O. Shwartzglass, and J. Shappir, Biosens. Bioelectron. **19**, 1703 (2004).
[21] S. Meyburg, M. Goryll, J. Moers, S. Ingebrandt, S. Bocker-Meffert, H. Luth, and A. Offenhausser, Biosens. Bioelectron. **21**, 1037 (2006).
[22] F. Heer, S. Hafizovic, W. Franks, A. Blau, C. Ziegler, and A. Hierlemann, IEEE J. Solid-State Circuits **41**, 1620 (2006).
[23] K. Imfeld, S. Neukom, A. Maccione, Y. Bornat, S. Martinoia, P. A. Farine, M. Koudelka-Hep, and L. Berdondini, Ieee Trans. Biomed. Eng. **55**, 2064 (2008).
[24] W. L. C. Rutten, Annu. Rev. Biomed. Eng. **4**, 407 (2002).
[25] A. Lambacher and P. Fromherz, Appl. Phys. A Mater. Sci. Process. **63**, 207 (1996).
[26] G. Wrobel, M. Höller, S. Ingebrandt, S. Dieluweit, F. Sommerhage, H. P. Bochem, and A. Offenhäusser, J. R. Soc. Interface **5**, 213 (2008).
[27] K. Toma, H. Kano, and A. Offenhäusser, ACS Nano **8**, 12612 (2014).
[28] M. E. Spira and A. Hai, Nat. Nanotechnol. **8**, 83 (2013).
[29] M. E. Spira, D. Kamber, A. Dormann, A. Cohen, C. Bartic, G. Borghs, J. P. M. Langedijk, S. Yitzchaik, K. Shabthai, and J. Shappir, Solid-State Sensors, Actuators Microsystems Conf. 2007. TRANSDUCERS 2007. Int. Source IEEE Xplore 1247 (2007).
[30] A. Hai, J. Shappir, and M. E. Spira, J. Neurophysiol. **104**, 559 (2010).

[31] A. Hai, D. Kamber, G. Malkinson, H. Erez, N. Mazurski, J. Shappir, and M. E. Spira, J. Neural Eng. **6**, 066009 (2009).
[32] A. Hai, A. Dormann, J. Shappir, S. Yitzchaik, C. Bartic, G. Borghs, J. P. M. Langedijk, and M. E. Spira, J. R. Soc. Interface **6**, 1153 (2009).
[33] D. Brüggemann, B. Wolfrum, V. Maybeck, Y. Mourzina, M. Jansen, and A. Offenhäusser, Nanotechnology **22**, 265104 (2011).
[34] D. Bruggemann, K. E. Michael, B. Wolfrum, and A. Offenhausser, Int. J. Nano Biomater. **4**, 108 (2012).
[35] G. Panaitov, S. Thiery, B. Hofmann, and A. Offenhäusser, Microelectron. Eng. **88**, 1840 (2011).
[36] F. Santoro, S. Dasgupta, J. Schnitker, T. Auth, E. Neumann, G. Panaitov, G. Gompper, and A. Offenhäusser, ACS Nano **8**, 6713 (2014).
[37] F. Santoro, E. Neumann, G. Panaitov, and A. Offenhäusser, Microelectron. Eng. **124**, 17 (2014).
[38] F. Santoro, J. Schnitker, G. Panaitov, and A. Offenhäusser, Nano Lett. **13**, 5379 (2013).
[39] C. Xie, Z. Lin, L. Hanson, Y. Cui, and B. Cui, Nat. Nanotechnol. **7**, 185 (2012).
[40] B. Hofmann, E. Kätelhön, M. Schottdorf, A. Offenhäusser, and B. Wolfrum, Lab Chip **11**, 1054 (2011).
[41] A. Czeschik, A. Offenhäusser, and B. Wolfrum, Phys. Status Solidi **211**, 1462 (2014).
[42] A. Czeschik, P. Rinklin, U. Derra, S. Ullmann, P. Holik, S. Steltenkamp, A. Offenhäusser, and B. Wolfrum, *Nanostructured Cavity Devices for Extracellular Stimulation of HL-1 Cells* (In Preparation, 2015).
[43] Z. C. Lin, C. Xie, Y. Osakada, Y. Cui, and B. Cui, Nat. Commun. **5**, 1 (2014).
[44] Y. Cui, Q. Wei, H. Park, and C. M. Lieber, Science **293**, 1289 (2001).
[45] F. Patolsky, B. P. Timko, G. Yu, Y. Fang, A. B. Greytak, G. Zheng, and C. M. Lieber, Science **313**, 1100 (2006).
[46] J. F. Eschermann, R. Stockmann, M. Hueske, X. T. Vu, S. Ingebrandt, and A. Offenhäusser, Appl. Phys. Lett. **95**, 083703 (2009).
[47] T.-S. Pui, A. Agarwal, F. Ye, N. Balasubramanian, and P. Chen, Small **5**, 208 (2009).
[48] E. Stern, J. F. Klemic, D. A. Routenberg, P. N. Wyrembak, D. B. Turner-Evans, A. D. Hamilton, D. A. LaVan, T. M. Fahmy, and M. A. Reed, Nature **445**, 519 (2007).
[49] X. T. Vu, J. F. Eschermann, R. Stockmann, R. GhoshMoulick, A. Offenhäusser, and S. Ingebrandt, Phys. Status Solidi **206**, 426 (2009).
[50] B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, and R. Thewes, IEEE J. Solid-State Circuits **38**, 2306 (2003).
[51] G. Zeck, A. Lambacher, and P. Fromherz, PLoS One **6**, e20810 (2011).
[52] E. Ferrea, A. Maccione, L. Medrihan, T. Nieus, D. Ghezzi, P. Baldelli, F. Benfenati, and L. Berdondini, Front. Neural Circuits **6**, 80 (2012).
[53] U. Frey, U. Egert, F. Heer, S. Hafizovic, and A. Hierlemann, Biosens. Bioelectron. **24**, 2191 (2009).
[54] D. J. Bakkum, U. Frey, M. Radivojevic, T. L. Russell, J. Müller, M. Fiscella, H. Takahashi, and A. Hierlemann, Nat. Commun. **4**, 2181 (2013).
[55] F. C. de Anda, G. Pollarolo, J. S. Da Silva, P. G. Camoletto, F. Feiguin, and C. G. Dotti, Nature **436**, 704 (2005).
[56] N. Arimura and K. Kaibuchi, Nat. Rev. Neurosci. **8**, 194 (2007).
[57] L. J. Richards, S. E. Koester, R. Tuttle, and D. D. M. O. Leary, J. Neurosci. **17**, 2445 (1997).

[58] M. Jungblut, W. Knoll, C. Thielemann, and M. Pottek, Biomed. Microdevices **11**, 1269 (2009).
[59] C. D. James, A. J. H. Spence, N. M. Dowell-Mesfin, R. J. Hussain, K. L. Smith, H. G. Craighead, M. S. Isaacson, W. Shain, and J. N. Turner, IEEE Trans. Biomed. Eng. **51**, 1640 (2004).
[60] S. Roth, G. Bugnicourt, M. Bisbal, S. Gory-Faure, J. Brocard, C. Villard, and S. Gory-Fauré, Small **8**, 671 (2012).
[61] R. Fricke, P. D. Zentis, L. T. Rajappa, B. Hofmann, M. Banzet, A. Offenhausser, and S. H. Meffert, Biomaterials **32**, 2070 (2011).
[62] A. J. Nichols, L. H. Carney, and E. C. Olson, BMC Neurosci. **9**, 50 (2008).
[63] B. T. Schaar and S. K. McConnell, Proc. Natl. Acad. Sci. U. S. A. **102**, 13652 (2005).
[64] M. Merz and P. Fromherz, Adv. Funct. Mater. **15**, 739 (2005).
[65] M. J. Jang and Y. Nam, J Neural Eng **9**, 46019 (2012).
[66] A. K. K. Vogt, G. J. J. Brewer, and A. Offenhäusser, Tissue Eng. **11**, 1757 (2005).
[67] A. K. Vogt, F. D. Stefani, A. Best, G. Nelles, A. Yasuda, W. Knoll, and A. Offenhausser, J Neurosci Methods **134**, 191 (2004).
[68] C. Xie, L. Hanson, W. Xie, Z. Lin, B. Cui, and Y. Cui, Nano Lett. **10**, 4020 (2010).
[69] O. Feinerman, M. Segal, and E. Moses, J Neurophysiol **94**, 3406 (2005).
[70] O. Feinerman, A. Rotem, and E. Moses, Nat. Phys. **4**, 967 (2008).
[71] J. M. Peyrin, B. Deleglise, L. Saias, M. Vignes, P. Gougis, S. Magnifico, S. Betuing, M. Pietri, J. Caboche, P. Vanhoutte, J. L. Viovy, and B. Brugg, Lab Chip **11**, 3663 (2011).
[72] B. Deleglise, B. Lassus, V. Soubeyre, A. Alleaume-Butaux, J. J. Hjorth, M. Vignes, B. Schneider, B. Brugg, J.-L. Viovy, and J.-M. Peyrin, PLoS One **8**, e71103 (2013).
[73] J. Albers, K. Toma, and A. Offenhäusser, Biotechnol. J. (2014).
[74] W. Li, Z. Xu, J. Huang, X. Lin, R. Luo, C.-H. Chen, and P. Shi, Sci. Rep. **4**, 4784 (2014).

INDEX

Index

Symbole

α-helix, I.14, E5.5
β-sheet, I.14, E5.7
2D blood model, F6.6
2D-NMR, E3.5
3D electrodes, G5.17

A

actin, I.14, F8.3
action potential (AP), A2.3, A5.8, G5.13
active Brownian particles, F7.17
active Brownian rods, F7.20
active Brownian spheres, F7.18
active Brownian spheres, collective motion, F7.22
active gels, I.22
active materials, I.21
active matter, I.21
adaptive materials, I.17
adenylate kinase, E3.8
adsorption, A4.7
advection-diffusion equation, B1.29
air retention, D3.3
alchemy, D3.3
algae, F7.8
Alzheimer's disease, E4.2
American Society for Testing and Material, D3.9
amide I absorption, E3.3
amino acids, A2.12
amphiphilic molecules, I.7
amyloid, E4.2
amyloid-β, E4.10
anticodon, A2.13
Aquaporin, I.15
ASTM, D3.9
atomic force microscopy (AFM), C4.3
atomic layer deposition (ALD), G5.11
atomic vibration, E3.3
ATP, A2.15
ATPase motor, E3.12
axoneme, F7.6
axons, A2.2

B

backbone dynamics, E3.8
backscattering spectroscopy, C1.27, G3.3
bacteria, F7.2
band gap, A1.6
barrier properties, D3.10
bending potential, B2.4
bending rigidity, D1.8, F8.7
bicontinuous microemulsion, D4.8
bimimetic synthesis, A3.4
binding induced protein folding, E2.11
bio-synthetic hybrid materials, I.21
biogenic silification, A3.5
biological barrier, F6.3
biological membranes, A5.2
biomaterial, A3.2
biomimetic microswimmers, F7.10
biomineralization, A3.3
biopolymer networks, I.14
biopolymers, I.14
Bjerrum-length, B4.14
block copolymer, A4.3
blood, F6.2
blood flow, F6.6
Boltzmann equation, B3.8
Boltzmann factor, B2.16
bond potential, B2.4
bone, A3.2
bottom-up approaches, G5.3
boundary condition, B3.19, F6.6
Brownian, A1.8
Brownian force, B1.9
Brownian motion, I.4
Brownian particles, active, F7.17
Brownian torque, B1.15

C

calcium carbonate (calcid), A3.3
calixarene, C4.9
canonical ensemble, B2.13
canonical thermostat, B3.12
capillary flow, C6.26
capillary interaction, D1.7
car tire, D3.3
carrier mechanism, G3.2

catalysis, A1.3
catch bonds, C4.12
cell adhesion, F6.3
cell-electrode cleft, G5.13
cellular signalling, A5.2
central limit theorem, I.9
chaperone, E2.13
charged spheres, B4.13
chemical shift, C5.8, E3.5
chemical vapor deposition (CVD), G5.9
Chlamydomonas reinhardtii, F7.8
cholesterol, A2.7
chromatids, A2.9
chromatin, I.13, A2.14
chromosomes, A2.9
cilia, F7.3
cilia, synchronization, F7.24
clay, D3.2
click chemistry, I.18
CMOS-type MEAs, G5.12
coherent scattering, C1.13
coiled-chain conformation, I.9
collagen, F8.5
colloid, B3.27, B4.2, D1.2, F1.4
colloidal crystals, D2.8
colloidal dispersion, B2.5
colloidal interaction, B4.2
colloidal lithography, D2.7
colloidal stability, B4.15
colloidal suspensions, I.5
colloidosome, D1.8
complex fluids, I.2
composite, D3.2
conductivity, G3.2
cone and plate geometry, C6.24
confocal microscopy, C3.4, F1.23, F2.3
conformational space, E2.4
conformational states, C3.9, E2.4
constraint release, G4.8
continuity equation: rotational, B1.13
continuity equation: translational, B1.8
contour length fluctuations, G4.7
contractions, B1.37
cooperative rotation, E3.12
core-shell, F5.5
correlation function, C1.34
corresponding states, B2.7
Couette geometry, C6.23
Coulomb potential, B2.5
creep compliance, C6.9
critical micelle concentration, D4.3
critical phenomena, I.5
crosslinked networks, F8.10
crystallography, E3.4
cytoskeleton, I.14, A2.5, D1.4
cytosol, A2.5

D

Debye-Hückel approximation, B4.11, F3.5
delta function, C1.32
demixing, B5.5
dendrites, A2.3
depletion attraction, I.7
depletion interaction, B4.17
Derjaguin approximation, B4.7
desorption, A4.7
desoxyribonucleic acid (DNA), I.13
detailed balance condition, B2.17
diatom, A3.2
diblock copolymer, I.13
differential cross-section, C1.3
diffraction, G3.3
diffraction limited resolution, C3.6
diffuse (double) layer, B4.13
diffusio-phoretic microswimmers, F7.9
diffusion, C1.17, G3.2
diffusion coefficient, B2.7, B3.17, F4.3
diffusive wave spectroscopy, F2.8
dipol-dipol coupling, C5.12
dispersion, D3.7
dispersion colloids, I.5
dissipative particle dynamics, B3.2, F6.13
dissociation kinetics, C4.3
DLVO interaction, B4.15
DNA, I.7, A0.13, A2.9
DNA origami, I.18
DNA-based materials, I.18
domain motions, E3.8
domains, E3.2
double differential cross-section, C1.11
DPD thermostat, F6.13
drug carrier, F6.3
drug delivery, F6.3
dynamic bonding, G1.3

dynamic light scattering (DLS), C1.24, F5.3
dynamic scattering, C1.11
dynamic structure factor, C1.13
dynamical mechanical analysis, D3.7

E

E coli, F7.2
Einstein's formula: rotational, B1.15
Einstein's formula: translational, B1.9
elastic normal modes, E3.10
elastic scattering, C1.3
elastic structure factor, C1.15
elastomeric stamps, D2.3
electric permittivity, B1.16
electric double layer, F3.2
electric double layer interactions, F3.10
electric field gradient, C5.10
electrochemical potential, A5.3
electrode polarization, F3.15
electrodeposition, G5.7
electrolyser, G2.3
electron beam lithography, D2.3
electronics, D3.2
electrophoretic mobility, F3.19
electrostatic interaction, B4.10
electrostatic pair potential, B4.14
electrostatic screening, B4.11
elongational flow, C6.27
encapsulation, G1.4
end-to-end distance, I.9
endoplasmic reticulum, A2.5
energy landscape, E2.4
energy transfer, C1.11
ensemble average, C3.2
entangled networks, F8.9
entropic elasticity, F8.8
epigenetic, A2.14
erythrocytes, A2.3
etching, G5.3
exfoliated, D3.10
extensional flow, C6.4, C6.8
extracellular signal, G5.13

F

fd-virus, I.6
fiber, hollow, B5.10
fibrin, F8.7
Fick's diffusion equation: rotational, B1.12
Fick's diffusion equation: translational, B1.5
field flow fractionation, F4.12
fieldeffect transistor (FET), G5.2
filtration, B5.2
finite size effects, B2.7
fish tail point, D4.15
flagella, F7.2
flippases, A2.9
Flory-Huggins model, B5.6
flow field, B3.27
fluid, shear-thinning, I.7
fluorescence, C2.10
fluorescence anisotropy, C3.12
fluorescence correlation spectroscopy (FCS), C3.10
fluorescence lifetime, C3.12
focussed ion beam (FIB), G5.19
folding funnel, E2.4
folding kinetics, E2.6
folding pathway, E2.3
folding/unfolding transition rate, E2.7
form factor, C1.10
Förster resonance energy transfer (FRET), C3.7
Fourier transform, C1.31
fractal, G3.7
free energy of an electric double layer, F3.11
free induction decay, C5.6
friction anisotropy, F7.15
fuel cell, G3.2
functional amyloid, E4.3

G

galvanization, G5.7
gas separation, B5.8
Gauss's integral theorem, B1.37
Gaussian approximation, C1.18
Gaussian curvature, D1.23, D4.8
Gaussian random fields theory, D4.20
Gaussian saddle splay modulus, D1.23
Gaussian, Markovian stochastic process, B3.4
gene, A2.10
generalized exponential function, G4.12
genome, A2.10

Gibbs phase triangle, D4.6
glass transition, D3.8
glutamate transporter, A5.5
glutamatergic synaptic transmission, A5.5
glycolipids, A2.7
glycosaminoglycan, C4.12
Golgi complexes, A2.6
Gouy-Chapman model, F3.2
grand challenges, I.24
graphene, G2.6
grazing incidence scattering, D4.22
Grotthus mechanism, G3.2
growing tissues, E1.1
Guinier approximation, C1.8

H

Hahn echo, C5.14
Hamaker constant, B4.6
hard spheres, B4.2
hard-core potential, I.6
head distortion temperature, D3.9
Helfrich free energy, D4.8
Helfrich interaction, D4.10
Helfrich model, D1.3
hematocrit, F6.4
heme, A2.3
heparansulfate, C4.12
hepatocytes, A2.4
hierarchical relaxation, G4.10
high pressure, F2.2
high pressure Microscopy, F2.10
hinge, E3.8
histones, A2.9
hollow microgels, F5.9
homeostatic pressure, E1.4
homopolymers, I.8
host-guest interaction, C4.8
HT-PEFC, G3.2
hydration, G2.10
hydration water, F2.4
hydrodynamic interactions, far-field approximation, F7.13
hydrodynamic synchronization, B3.24, F7.23
hydrodynamics, B3.3
hydrodynamics interactions, squirmers, F7.14
hydrodynamics, dipole near wall, F7.29
hydrodynamics, dipole tensor, F7.29
hydrodynamics, Oseen tensor, B3.7, F7.28
hydrogen bond, E5.2
hydrogen bonding, I.19, G1.6
hydrophilic, I.7
hydrophobic, I.7
hydrophobic collapse, E2.3
hydroxyapatite (crystals), A3.2

I

impermeability, D3.3
incoherent scattering, C1.12
inelastic scattering, C1.11
inelastic x-ray scattering, C1.26
infinity optics microscope, C2.3
inorganic particle, D3.2
intercalated, D3.10
interface, A1.2, D1.2
intermediate filaments, F8.4
intermediate scattering function, C1.13
International Organization for Standardization, D3.9
intracellular membrane systems, A5.7
intrinsically disordered proteins (IDP), E2.11, E3.11
inverse modelling, G4.12
inversion recovery experiment, C5.20
ion condensation, F3.7
ion diffusion, A5.2
ionomer, G3.2
ISO, D3.9
isobaric-isothermal ensemble, B2.14

K

kinesin, I.21
Kramer's diffusion theory, E2.7
Kramers-Bell-Evans theory (KBE-theory), C4.3
Kuhn, F1.18

L

LaMer, A1.11
Landau expansion, D4.11
Landau-Lifshitz Navier-Stokes equation, B3.5
Langevin, A1.8
Langevin equation, F7.15
Langmuir-Blodgett film, A4.2

large amplitude oscillatory shear flow, F1.20
large scale dynamics., E3.12
Larmor frequency, C5.3
laser, D3.3
lattice Boltzmann, B3.2
law of large numbers, I.9
layer deposition, G5.5
Lennard Jones potential, B2.4
leukocytes, A2.3
lift force, F6.3
light microscope, condenser, C2.5
light scattering, C1.23
line shape analysis, C5.15
linear viscoelasticity, C6.9
lipid, I.16, B2.5
lipid bilayer, I.16, D1.4
lipid rafts, A2.9
liquid, C1.9
liquid crystal, A4.5
liquid crystal display, D3.3
lithography, D2.2, G5.3
living polymers, I.13
locomotion, F7.2
long-chain branching, G4.11
long-range hydrodynamic interactions, B3.16
long-time tail, B3.7, B3.15
loss modulus, C6.11, C6.16
Lotus leaf, I.20
low-Reynolds-number hydrodynamics, F7.27
lyotropic phases, I.8, D4.5
lysosomes, A2.6

M

magic angle spinning, C5.23
magnetic heating, A1.7
magneto-rheological fluid, I.18
malaria parasite, D1.4, D1.15
mammalian central nervous system, A5.5
margination, F6.7
margination diagram, F6.8
margination probability, F6.8
material functions, C6.3
Maya, D3.2
mean curvature, D1.23, D4.8
mean-square displacement, B1.7, B6.21, C1.18
membrane, B5.2, D1.2, D1.4, D1.8
membrane fabrication, B5.9
membrane lipid, I.16
membrane potential, A5.2
membrane transport proteins, A5.2
membrane, asymmetric, B5.3
mesoscale simulation, B3.2
mesoscopic method, F6.13
metachronal wave, F7.3, F7.24
metachronal wave, transport properties, F7.25
Metropolis algorithm, B2.17
micelle, I.8, D4.3
micro-contact printing, D2.3
micro-organisms, F7.2
microcanonical ensemble, B2.10, B2.13
microcomposites, D3.10
microcontact printing, G5.2
microelectrode array (MEA), G5.2
microemulsions, I.8
microfluidic, B2.14
microgel, A4.8, F5.2
micropillars, D2.3
microrotor, B3.24
microscale thermophoresis (MST), F4.15
microswimmer, I.22, B1.19, F7.2
microswimmer, biomimetic, F7.8
microswimmer, chiral, F7.17
microswimmer, diffusio-phoretic, F7.9
microswimmer, diffusion coefficient, F7.18
microswimmer, force dipole, F7.13
microswimmer, puller, F7.13
microswimmer, pusher, F7.13
microswimmer, rotator, F7.11
microswimmer, surface aggregation, F7.18
microswimmer, synthetic, F7.8
microswimmer, thermo-phoretic, F7.9
microswimmer, three-bead, F7.10
microswimmers, surface accumulation, F7.20
microtubules, I.15, F8.4
microvascular flow, F6.8
minimum image convention, B2.9
mitochondrion, A2.6
mixture, D3.2

molecular chaos, B3.8, B3.18
molecular dynamics, B2.1, E4.10
molecular weight, C1.23
Monge representation, D1.21
monolayer, self-assembled, A4.2
monomer, I.8
Monte Carlo, B2.1, B2.16
Montmorillonite, D3.3
motional narrowing, C5.17
motor neurons, A2.2
motor protein, I.21
multi-domain protein folding, E2.14
multiparameter burst analysis, C3.15
multiparticle collision dynamics, B3.2
myosin, I.21

N

nacre, A3.4
Nafion, G3.2
nano imprint lithography, D2.3
nanobiomaterial, E4.7
nanocavities, G5.21
nanocomposite, D3.3
nanoparticle, A3.6, D1.2, D1.8
nanopillar, G5.18
nanopropeller, I.22
nanorods, bimetallic, F7.9
nanoscale, D3.3
nanotechnology, G5.2
Navier-Stokes equation, B1.33, B3.2, B3.3, F7.12
Néel remagnetization, A1.8
nematic, F1.13
nematic-isotropic phase transitions, F2.12
Nernst-Einstein equation, G3.13
neuronal networks, G5.25
neurons, A2.2
neutron scattering, C1.27
neutron spin echo spectroscopy, C1.28, E3.9, G3.5
Newton's second law, F6.13
normal mode deformations, E3.10
nuclear electric quadrupole moment, C5.7
nuclear magnetic dipole moment, C5.7
nuclear magnetic resonance (NMR), E3.5, F1.9
nucleation and growth, F2.8
nucleosomes, A2.9
nucleus, A2.6

O

Onsager coefficient, F4.16
Opalina, F7.3
optical coherent tomography, F1.12
Oseen tensor (hydrodynamics), B3.7, F7.28
overdamped dynamics, B1.3

P

packing parameter, D4.4
paint remover, D4.21
pair correlation function, B2.6, C1.5
parallel plate geometry, C6.24
Paramecium, F7.3
particle, D1.2, D3.2
particle distribution, F6.7
particle image velocimetry, F1.8
partition function, B2.16
Péclet number, F7.18
Peltier effect, F4.4
percolation threshold, F2.9
periodic boundary conditions, B2.8
persistence length, I.15, F8.7
pervaporation, B5.8
phase prism, D4.13
phase problem, E5.13
phase segregation, A4.4
phase separation, B5.5
phase transitions, I.5
PhoB, C4.7
phosphatidyl-choline (PC), I.16
phosphoglycerate kinase, E3.9
Phospholipids, A2.7
phosphoric acid, G3.2
phosphorylation, E5.10
photolithography, D2.2
photon correlation spectroscopy, C1.24, F1.6
photoresist, D2.2
photoswitch, C4.9
physical vapor deposition (PVD), G5.6
Pickering emulsion, I.17
planar microelectrode arrays, G5.14
plasma membrane, A2.4
plasmonic nanoparticles, A1.4
plumbers nightmare, D4.5
point charges, B4.10

point spread function, C2.7
Poiseuille flow, B2.14, B3.23
Poisson equation, B4.10
Poisson-Boltzmann equation, B4.11, F3.5
polar filaments, I.15
polarization, F1.13
polybenzimidazole, G3.2
polyelectrolytes, F5.7
polymer, B2.5, D3.2, F1.4
polymer architecture, I.8, G4.2
polymer boosting effect, D4.16
polymer brush, B4.22
polymer dynamics, G3.3
polymer electrolyte, G3.2
polymer entanglements, G4.3
polymer entropy, I.9
polymer membrane, G2.4
polymer-based solar cells, I.21
polymeric additives, I.10
polymerization, F5.2
pom-pom molecule, G4.10
Porod law, C1.9, G3.8
potassium channels and chloride channels, A5.8
precipitation, F5.2
preparation of nanocomposites, D3.6
presynaptic terminal, A2.3
primitive path, G4.3
probability density functions, B1.3
promoter, A2.15
protease ClpP, E3.7
protein, I.14
protein aggregation, E4.3
protein data bank (PDB), E5.11
protein DNA interaction, C4.7
protein dynamics, C3.9
protein folding, E2.2
protein misfolding, E2.2
protein regulated mineralization, A3.6
protein secondary structure, E2.3
protein stability, E2.6
protein unfolding, C3.12
protein, primary structure, I.14
protein, secondary structure, I.14
protein, tertiary structure, I.14
proton conducting membrane, G3.2
proton conductivity, G2.7
puller (microswimmer), F7.13
pulsed interleaved excitation (PIE), C3.16
pusher (microswimmer), F7.13

Q

quadrupolar interaction, C5.10
quantum confinement effect, A1.6
quasielastic neutron scattering (QENS), G3.3
quasielastic scattering, C1.18

R

radius of gyration, C1.8
random walk, persistent, F7.18
ray optics of the light microscope, C2.2
RBC model, F6.4
RBC-free layer, F6.7
reactive ion etching (RIE), G5.5
receptor-ligand interactions, C4.6
recognition forces, C4.3
red blood cell, I.16, F6.2
reduced units, B2.7
refinement, E5.15
relaxation modulus, C6.10, C6.16, G4.3
relaxation spectrum, C6.16, C6.32
renormalization, D4.15
replication, A2.9
reptation, F8.9, G4.4
resistive-force theory, F7.16
resolution of the light microscope, C2.7
responsive materials, I.17
Reynolds number, B3.4, F7.12
rheology, C6.3, F1.2, G4.11
rheology of tissues, E1.8
ribosomes, A2.5
RNA, A2.9
rod-like colloids, B2.5
rods, self-propelled, F7.20
rotating frame, C5.5
rotational diffusion, B1.10
rotational protein diffusion, C3.15
rough endoplasmic reticulum, A2.6
Rouse model, B6.40
Rouse relaxation, G4.4
rubber elasticity, I.9
rubbers, B6.3
run-and-tumble motion, F7.2

S
saddle splay modulus, D1.8
Salmonella, F7.3
saturated monolayer, I.8
scallop theorem, B3.4, F7.13
scattering, C1.2
scattering function, C1.13
scattering length, C1.3
scattering length density, C1.5
scattering techniques, F2.3
scattering vector, C1.3
Schmidt number, B3.19
scratch resistant coating, D3.3
screening length, B4.11, B4.14
seal resistance, G5.14
secondary structure, E5.5
sedimentation, B3.27
Seebeck effect, F4.4
self correlation function, B6.41
self-cleaning materials, I.20
self-healing materials, I.19, G1.3
self-propelled particle, I.22, F7.2
self-propelled rods, F7.20
self-propelled spheres, F7.18
semiflexible polymers, F8.8
sensor, D3.3
shear flow, B1.22, B2.15, B3.22, C6.3, C6.22
shear modulus, I.3
shear thickening, C6.6
shear thinning, C6.5, C6.19
shear-banding, B1.29
shear-curvature viscosity, B1.33
shear-thinning fluid, I.7
shot-noise limit, C3.9
sidechain movements, E3.4
silaffins, A3.5
silica, A3.4
silicon nanowire (SiNW), G5.22
simulations, F8.12
single molecule, C3.2
single molecule force spectroscopy (SMFS), C4.2
single molecule techniques, E2.9
single-domain, A1.7
sliding plate rheometer, C6.22
slip bonds, C4.12
small angle light scattering, F1.16
small-angle neutron scattering (SANS), C1.27, D3.6, F1.18, F5.3, G3.3
small-angle x-ray scattering (SAXS), C1.26, F1.20
Smoluchowski equation, B1.24
soccer ball, D3.3
sodium dodecyl sulfonate (SDS), I.7
soft lithography, D2.3
soft matter, I.2, B3.2
soil remediation, D4.25
solar cells, I.21
soma, A2.3
Soret effect, F4.2
spectral density, C5.18
sperm, F7.3
sperm, chiral, F7.17
sperm, swimming velocity, F7.17
spheres, self-propelled, F7.18
spherical form factor, C1.7
spin interaction, C5.7
spin lattice relaxation, C5.6
spin population, C5.4
spin precession, C5.4
spin spin relaxation, C5.18
spinodal decomposition, F2.8
spontaneous curvature, D1.9, D1.23
squirmer, F7.10
squirmer interactions, F7.14
star chains, G4.13
static scattering, C1.3
statistical polymer composition, G4.13
steric repulsion, B4.21
Stern layer, F3.2
stick-slip boundary conditions, B2.14
stimulated echo, C5.20
stochastic process, B3.4
stochastic rotation dynamics, B3.11
Stokes equation, B3.4, F7.12
Stokes's integral theorem, B1.37
storage modulus, C6.11, C6.16
Stoßzahlansatz, B3.8
strain-stiffening, F8.11
stress tensor, B3.3
structure factor, B3.27, C1.5, E5.12
structure of channel proteins, A5.4
subpopulation, C3.2

sulfatase, C4.12
sum frequency generation, A3.6
super-hydrophobic surface, I.20
superellipsoid, D1.4
supramolecular chemistry, G1.7
Surface Evolver, D1.24
surface forces apparatus, B4.23
surface freezing, A4.5
surface induced ordering, A4.2
surface plasmon resonance, A1.4
surfactant, I.7, D4.2, F1.3
suspensions, I.5
swarming, F7.11
swimming, helical, F7.17
synchronization, hydrodynamic, F7.23

T

tactoid, I.6
Taylor dispersion, B1.22
temperature gradient, B2.15, F4.2
temperature sensitive materials, F5.8
Teubner-Strey model, D4.19, G3.8
thermal diffusion coefficient, F4.3
thermal diffusivity, F4.3
thermal oxidation, G5.8
thermo-phoretic microswimmers, F7.9
thermo-responsive, F5.6
thermodynamic hypothesis in protein folding, E2.5
thermodynamic limit, B2.7
thermophoresis, F4.2
time correlated single photon counting (TCSPC), C3.5
time resolved, E3.4
time resolved anisotropy decay, C3.14
time-of-flight spectroscopy, C1.27, G3.4
TIRF microscope, C3.3
tissue formation, A3.2
tissue spheroids, E1.10
tobacco mosaic virus (TMV), I.7
top-down approaches, G5.3
torsion potential, B2.4
total internal reflection microscopy, B4.24
transcription, A2.12
transient network, I.19
transition network, E4.10
transition state theory (Eyring equation), E2.6
translational diffusion, B1.4
translational protein diffusion, C3.12
transmission electron microscopy (TEM), G3.5
transmitters, A2.3
transport coefficient, B3.17
transport mechanism, A5.3
transversal magnetization, C5.6
triangulated surfaces, D1.4
tube diameter, B6.22
tube model, G4.6
two dimensional NMR, C5.20

U

ultra-sound velocimetry, F1.9
universality, I.9, A0.11

V

van der Waals interaction, B4.3
van Hove correlation function, C1.15
vapour permeation, B5.8
velocity correlation function, B3.6, B3.15
Vicsek model, F7.11
virus, D1.15
viscoelasticity, F8.9
viscosity, B3.18
voltage-gated sodium channels, A5.8
volume phase transition temperature, F5.4
Volvox, F7.8

W

water, G2.7
white blood cell, F6.3
Wiener-Khintchine theorem, C1.2
wormlike micelles, F1.3
wrapping phase diagram, D1.10
wrinkling, D2.10

X

x-ray crystallography, E5.11
x-ray photon correlation spectroscopy, C1.26, F1.23
x-ray scattering, C1.25, D3.4, F8.12

Y

yield stress, B1.33, C6.6

The IFF Spring School & Solid-State, Soft-Matter and Biophysics Research in Jülich

PGI-JCNS-TA
52425 Jülich
Phone: *++49 2461 61-1739*
Fax: *++49 2461 61-2410*
E-Mail: *springschool@fz-juelich.de*
web: *www.iff-springschool.de*

The annual IFF Spring School arises from the tradition of the Institut für Festkörperforschung (IFF) which was founded in 1969. The institute's research topics ranged from electronic and structural properties of solids, and nanoelectronics, to the thermal and dynamical behavior of soft matter. The IFF had organized this spring school for over 40 years. Since the restructuring in 2011, research of electronic systems and phenomena, as well as their applications in information technology, is combined in the Peter Grünberg Institut (PGI) named after the IFF scientist who received the Nobel Prize for physics in 2007. Soft matter and biophysics research is integrated in the Institute of Complex Systems (ICS). These institutes are linked together and supported by the Institute for Advanced Simulation (IAS), which focuses on developing and applying high-performance computing to understand complex systems, and the Jülich Centre for Neutron Science (JCNS), which is dedicated to the operation of neutron scattering instruments and national and international neutron sources. The IFF Spring School is now organized in turns by PGI, ICS and JCNS.

The Institute of Complex Systems (ICS) consists of 8 departments: neutron scattering, theoretical soft matter and biophysics, soft matter, cellular biophysics, molecular biophysics, structural biochemistry, biomechanics, and bioelectronics. Designing functional soft materials, predicting their properties and fine- tuning their performance is the visionary goal of soft matter science. A major objective of biophysics is to understand processes far from equilibrium, which distinguish dead and living matter. Proteins are exquisitely designed nano-machines; understanding their structural and dynamical properties on the basis of elementary molecular interactions and physical principles is a prerequisite for the development of cures for many diseases.

An essential part of the mission of the Institute of Complex Systems is the interdisciplinary education of graduate students at the interface between physics, chemistry and biology. Here, the International Helmholtz Research School on Biophysics and Soft Matter (IHRS BioSoft) provides a program of introductory and advanced lectures, seminars, lab courses, retreats, and transferable skills courses. More information is provided on the webpage http://www.ihrs-biosoft.de.

Further the institute is coordinating the Marie Skłodowska Curie Initial training network SOMATAI which provides continuous scientific education and soft skills training to a team of 14 early stage researchers. Further details are given at
https://somatai.eu.

Forschungszentrum Jülich - Campus Map



Band / Volume 81
Interdomain Functional Dynamics of Phosphoglycerate Kinase Studied by Single-Molecule FRET
M. Gabba (2014), v, 179 pp
ISBN: 978-3-89336-943-0

Band / Volume 82
Silizium Nanoribbon Feld-Effekt Transistoren zur Kopplung an elektroaktive Zellen
M. Jansen (2014), xvi, 181 pp
ISBN: 978-3-89336-944-7

Band / Volume 83
Microscopic description of the inverse Faraday effect at subpicosecond time scales
D. Popova (2014), 183 pp
ISBN: 978-3-89336-962-1

Band / Volume 84
Neutron Scattering
Lectures of the JCNS Laboratory Course held at Forschungszentrum Jülich and at the Heinz Maier-Leibnitz Zentrum Garching
edited by Th. Brückel, G. Heger, D. Richter, G. Roth and R. Zorn (2014),
ca. 360 pp
ISBN: 978-3-89336-965-2

Band / Volume 85
Neutron Scattering
Experiment Manuals of the JCNS Laboratory Course held at Forschungszentrum Jülich and at the Heinz Maier-Leibnitz Zentrum Garching
edited by Th. Brückel, G. Heger, D. Richter, G. Roth and R. Zorn (2014),
ca. 195 pp
ISBN: 978-3-89336-966-9

Band / Volume 86
Development and Characterization of a Microfluidic Magnetic Oscillation Reactor for Enzymes
D. Jussen (2014), xxi, 131 pp
ISBN: 978-3-89336-974-4

Band / Volume 87
Submolecular imaging with single particle atomic force sensors
G. Kichin (2014), 140 pp
ISBN: 978-3-89336-976-8

Band / Volume 88
Multiscale Multimodel Simulation of Micromagnetic Singularities
C. Andreas (2014), xix, 188 pp
ISBN: 978-3-89336-983-6

Band / Volume 89
***Ab initio* description of transverse transport due to impurity scattering in transition-metals**
B. C. Zimmermann (2014), 164 pp
ISBN: 978-3-89336-985-0

Band / Volume 90
Ladungstransport durch Graphenschichten und GaAs-Nanodrähte untersucht mit einem Multispitzen-Rastertunnelmikroskop
S. Korte (2014), 96 pp
ISBN: 978-3-89336-990-4

Band / Volume 91
6th Georgian-German School and Workshop in Basic Science
A. Kacharava (Ed.) (2014), CD
ISBN: 978-3-89336-991-1

Band / Volume 92
Ab initio investigations of π-conjugated-molecule-metal interfaces for molecular electronics and spintronics
M. Callsen (2014), viii, 155 pp
ISBN: 978-3-89336-992-8

Band / Volume 93
Ladungstransportmessungen an Si(111) Oberflächen mit einem Multispitzen-Rastertunnelmikroskop
M. Blab (2014), iv, 132, X pp
ISBN: 978-3-89336-997-3

Band / Volume 94
Functional Soft Matter
Lecture Notes of the 46th IFF Spring School 2015
23 February – 06 March, 2015 Jülich, Germany
ed. by J. Dhont, G. Gompper, G. Meier, D. Richter, G. Vliegenthart, R. Zorn
(2015), ca. 600 pp
ISBN: 978-3-89336-999-7

Weitere ***Schriften des Verlags im Forschungszentrum Jülich*** unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp